

P25591 Ph



2010 M&I. 2 6

A Magyar Tudományos Akadémia  
Filozófiai Bizottságának folyóirata

# MAGYAR FILOZÓFIAI SZEMLE

2010/1 (54. évfolyam)

Filozófiai  
naturalizmus



## A SZERKESZTŐSÉG

Schmal Dániel felelős szerkesztő

Ambrus Gergely

Mekis Péter

Mester Béla

Ullmann Tamás

Valastyán Tamás

## A MAGYAR FILOZÓFIAI SZEMLE SZERKESZTŐBIZOTTSÁGA

Forrai Gábor egyetemi tanár (ME), a szerkesztőbizottság elnöke

Almási Miklós, az MTA rendes tagja

Altrichter Ferenc associate professor (University of North Carolina at Wilmington)

Csejtei Dezső egyetemi tanár (SZTE)

Fehér Márta egyetemi tanár (BMGE)

Heller Ágnes, az MTA rendes tagja

Hévízi Ottó tudományos tanácsadó (MTA Filozófiai Intézet, Lukács Archivum)

Hont István, fellow (King's College, Cambridge)

Kelemen János, az MTA levelező tagja (ELTE)

Klima Gyula egyetemi tanár (Fordham University, New York)

Máté András egyetemi docens (ELTE)

Rózsa Erzsébet egyetemi tanár (DTE)

Szabó Zoltán egyetemi tanár (Yale)

Tengelyi László egyetemi tanár (Wuppertal)

Weiss János egyetemi tanár (PTE)

A szerkesztőség elektronikus postacíme:

[filszemle@webmail.phil-inst.hu](mailto:filszemle@webmail.phil-inst.hu)

ISSN 0025-0090

Kiadja a Gondolat Kiadó

Felelős kiadó Bácskai István

1088 Budapest, Szentkirályi u. 16.

Telefon: +361-486-1527

[www.gondolatkiado.hu](http://www.gondolatkiado.hu)

<http://gondolatkiado.blog.hu>

Tördelő: Lipót Éva

Lapterv: Andorka Tímea és Pintér László

Nyomás és kötés: ETO-Print Kft.

A lap megjelenését a Nemzeti Kulturális Alap  
és a Magyar Tudományos Akadémia Filozófiai Bizottsága támogotta



2010 MAI. 26

# MAGYAR FILOZÓFIAI SZEMLE

2010/1 (54. évfolyam)

A Magyar Tudományos Akadémia  
Filozófiai Bizottságának folyóirata

Filozófiai  
naturalizmus



# Tartalom

Elöljáróban	4
-------------	---

## TANULMÁNYOK

Filozófiai naturalizmus (Ambrus Gergely)	5
TENGELYI LÁSZLÓ: A fenomenológia módszertani transzcendentalizmusa	9
HUORANSZKI FERENC: Filozófiai naturalizmus és természeti fajták	23
SIVADÓ ÁKOS: Fajtáink eredete	46
SZALAI MIKLÓS: Naturalizmus, szkepticizmus és racionalitás – Plantinga evolúciós szkeptikus érve	59
PLÉH CSABA: Naturalizmus a pszichológiában?	79
ZEMPLÉN GÁBOR – KUTROVÁTZ GÁBOR: A tudomány heterogenitása és a naturalizmus	87

## FÓRUM

LAKI JÁNOS: Tapasztalat, nyelv, társadalom, tudományfejlődés – Kuhn még egyszer	113
PLÉH CSABA: Laki János – A tudomány természete. Thomas Kuhn és a tudományfilozófia történeti fordulata	122
ZEMPLÉN GÁBOR: Kuhn recepciója, Laki rekonstrukciója	128
FORRAI GÁBOR: Laki, Kuhn és a realizmus	136
SCHWENDTNER TIBOR: A hermeneutika helye Thomas Kuhn gondolkodásában. Megjegyzések Laki János Kuhn-értelmezéséhez	148
LAKI JÁNOS: Hogyan konstruáljuk a szerzőt és a világot?	155

## SZEMLE

MÁRTON MIKLÓS: Philosophia perennis (Ujvári Márta könyvéről)	179
KLEIN DÁVID: Metafizikai hullámvasút (Tózsér János könyvéről)	185
ANGOL NYELVŰ ÖSSZEFOGLALÓK	194

# Előjáróban

A Magyar Tudományos Akadémia Filozófiai Bizottsága 2010-től új szerkesztőséget és szerkesztőbizottságot bízott meg a *Magyar Filozófiai Szemle* körüli teendők ellátásával. A folyóirat nemcsak külső és belső formájában újult meg: az idei első lapszám egyszersmind az új szerkesztőség első kísérlete is arra, hogy tematikus blokkban járjon körül egy témát. A lap évente három alkalommal jelenik meg, a rendes lapszámokhoz pedig időről időre magyar vagy idegen nyelvű különszámok csatlakoznak.

A *Magyar Filozófiai Szemle* az idei évtől referált folyóiratként működik. Mindez reményünk szerint a visszajelzéseknek köszönhetően a tárgyyszerű szakmai viták katalizátoraként is működik majd. *Minden* beérkezett tanulmányt anonim bírálóknak adunk ki elolvasásra. A bírálók javaslatokat tesznek az írások közlésével vagy elutasításával kapcsolatban, illetve – szükség esetén – változtatásokat javasolnak a közlésre szánt szövegben. A szerkesztőség a bírálók javaslatát tiszteletben tartja, de fenntartja magának a jogot, hogy a beérkezett szövegekkel kapcsolatban saját, szerkesztői szempontjait is érvényesítse.

Az anonim referálás fáradtságos munka. Ezúton köszönjük mindazoknak, akik a jelen lapszám elkészítésében a segítségünkre voltak, s gyakran igen rövid idő alatt készítettek színvonalas szakmai bírálatokat a beérkezett tanulmányokról. Köszönettel tartozunk továbbá Demeter Tamásnak, aki a *Fórum* rovatot szerkesztette. A *Filozófiai naturalizmus* című tematikus blokk összeállítása Ambrus Gergely munkája.

A *Magyar Filozófiai Szemle* az egész gazdag és sokszínű hazai filozófiai élet fóruma kíván lenni. Ezért kérjük és biztatjuk olvasóinkat, hogy – tematikától függetlenül, vagy a tematikus felhívásokhoz kapcsolódva – juttassák el hozzánk közlésre szánt tanulmányaikat és recenzióikat, valamint osszák meg velünk a lapszámmal vagy az egyes tanulmányokkal kapcsolatos észrevételeiket. Elektronikus postacímünk: [fil szemle@webmail.phil-inst.hu](mailto:fil szemle@webmail.phil-inst.hu)

*A szerkesztőség*

## Filozófiai naturalizmus

A „naturalizmus” a 20. századi és a kortárs filozófia egyik kitüntetett kérdése. Természetesen, amint az a filozófiában gyakran történni szokott, magának a fogalomnak a meghatározása is vita tárgya. Kiindulópontként talán használható az a hagyományos felfogás, amely módszertani és ontológiai naturalizmust különböztet meg. Az előbbi szerint a tudományok módszere episztemikusan felsőbbrendű minden más megismerő módszernél. Az utóbbi szerint az ontológia szempontjából is a tudományoké a végső szó: a legjobb tudományos elméletek mondják meg, mi létezik. Persze, ha ezt elfogadjuk, akkor felmerül a kérdés: mit tesz hozzá a filozófia a tudományokhoz? Nos, a naturalisták mondhatják: a tudományok és a filozófia *folytonos* egymással, a filozófiai érvelés és a fogalom-elemzés, illetve az empirikus kutatások nem választhatók el élesen, egymást kiegészítő tevékenységek. Másfelől, egy tartalmasabb ontológiai naturalizmusnak alapvető feltevése az is, hogy léteznek *természeti fajták*, tehát hogy fajtafogalmaink legalább egy része *nem* szociálisan konstruált. Ezt a feltevést (sokak szerint) a tudományok önmagukban nem döntenek el, önálló filozófiai kutatás tárgya. Természetesen – amint az alábbi tanulmányokból is kiderül – a naturalizmus e kiinduló értelmezési keretének a helytállósága vagy akár értelmessége is vitatható.

E számunk tanulmányai részben a naturalizmussal kapcsolatos kortárs viták egy-egy szűkebb témáját tárgyalják, részben pedig általánosabban, szélesebb történeti kontextusban közelítik meg a kérdést. Bár a filozófiai naturalizmus körüli viták ma elsősorban az analitikus metafizikában és a tudományfilozófiában zajlanak (kisebb mértékben a társadalomfilozófiában és az etikában), mindazonáltal a naturalizmushoz való viszony a 20. századi filozófia egész története szempontjából is fontos orientáló pont. Elegendő a „tudományos filozófia” mibenléte körül a megelőző századfordulón folytatott vitákra, vagy a logika és a matematika pszichologista és naturalista értelmezésének husserli és fregei elutasítására gondolni ahhoz, hogy Husserl transzcendentális fordulatának vizsgálata, amely

Tengelyi László nyitótanulmányának a középpontjában áll, kitűnő kiindulópontnak tűnjön a naturalizmus problémájának vizsgálata során.<sup>1</sup>

Tengelyi a *Husserliana* közelmúltban publikált két kötetének szövegei alapján Husserl transzcendentális fordulatának részleteit vizsgálja. Egyfelől rekonstruálja Husserl egy olyan érvét a transzcendentális idealizmus mellett, amely független az *Eszmék*ből ismert érveléstől (amely a tiszta tudat és a tér-időbeli dolgok régiójának elkülönültségét a dolgok esetleges és az élmények szükségszerű léte közötti különbségére alapozza). Másfelől olyan husserli gondolatokat körvonalaz, amelyek a transzcendentális idealizmus helyett egy módszertani transzcendentális felé mutatnak. Ez egy olyan „metafizikailag semleges” nézőpontot tesz lehetővé, amely – Tengelyi érvelése szerint – egyidejűleg haladja meg a naturalizmust és a szubjektivizmust, avagy a realizmust és az idealizmust.

Hasonlóan átfogó – bár modernebb tudományfilozófiai belátásokra támaszkodó – kritikát fogalmaz meg a naturalizmussal szemben a tanulmányblokk végén Zemplén Gábor és Kutrovács Gábor. Tézisük szerint „a tudományok elemzése és empirikus vizsgálata az utóbbi évtizedekben komolyan megkérdőjelezte annak a tudományképnek a tarthatóságát, amely a filozófiai naturalizmus legtöbb típusának háttérében áll. Ez pedig azt jelenti, hogy a naturalista megközelítések egy jelentős része, miközben a tudományos válaszok primátusát hirdeti, éppen a saját előfeltevéseinek tudományos megalapozatlanságát nem ismeri fel.” Más szóval, a tudományok heterogenitása nem fér össze sem a filozófiai naturalizmus módszertani, sem ontológiai alapfeltevéseivel. Mindazonáltal a szerzők továbbra is fenntartják a tudomány és a filozófia folytonosságának naturalista tézisét, és úgy vélik: a filozófiának igenis informálnak kell lennie releváns tudományos kérdésekben, a tudomány számára pedig a filozófia (és ezen belül a tudománytanulmányok) szolgálhatnak eszközökkel a tudományok önmegértése, a tudományos tudás természetének vizsgálata számára.

Egy ilyen – a pszichológiában és a kognitív tudományban évtizedek óta hatékony – együttműködés tapasztalatai Pléh Csaba tanulmányában egy az eddigiekkel ellentétes konklúziókhöz vezetnek. Pléh (sok természettudományos orientációjú pszichológussal ellentétben) nemcsak a módszertani, hanem az ontológiai naturalizmust is vállalja; ugyanakkor, mint fogalmaz, naturalizmusa „szemérmes”, amennyiben csak annyit tesz fel, hogy a tudat és az agy kapcsolata tudományos kérdés, amely tudományos választ fog kapni, ennek részleteire vonatkozóan azonban nincsenek előzetes filozófiai megszorításai. E megközelítés

<sup>1</sup> Másfelől azt is érdemes felidézni, hogy nemcsak a fenomenológiai hagyomány, hanem az analitikus filozófia fontos irányzatai és képviselői sem voltak naturalisták a 20. század első felében: sem a korai cambridge-i analitikusok, Russell, Moore és Wittgenstein a századelőn, sem a hétköznapi nyelv-filozófiai irányzat nem volt naturalista, és a hivatalos logikai pozitivisták doktrína is csak módszertani értelemben volt az. A quine-i fordulat előtti analitikus filozófiát naturalistának minősíteni tehát legalábbis félrevezető.



naturalizmusa pozitíve abban nyilvánul meg, hogy „az emberi természet” fogalmára épít, s az elmét biológiai szerveződésként fogja fel. Ugyanakkor a biológiai meghatározottság – konkrétan az a nézet, hogy az emberi természet tág értelemben genetikailag meghatározott – nem jelenti a determinizmus elfogadását. A fő riválissal, az elme szociális konstruktivista megközelítésével szemben ez az újfajta naturalista pszichológia a társasság elemi formáit is az emberi természethez sorolja, s a kultúrát olyan evolúciósan kialakult, „beállító tényezőként” értelmezi, amely nem áll szemben a természettel.

A tanulmányblokk további írásai szűkebb témaköröket vizsgálnak, ugyanakkor szoros kapcsolatban állnak a naturalizmus mibenlétének és védhetőségének a problémáival. Különösen igaz ez Szalai Miklós írására, aki Alvin Plantinga híres antinaturalista érvét elemzi. Plantinga szerint a naturalizmus inkohereus, öncáfoló állítás: ha ugyanis megismerő képességeinket az evolúció alakította ki, akkor általában kicsi a valószínűsége annak, hogy hiteink a világról igazak, beleértve magát az evolúcióba vetett hitünket is. Más szóval: aki elfogadja a naturalizmust és azt, hogy kognitív képességeink evolúciós eredetűek, az egyszerűen egy cáfoló megfontolással is rendelkezik a naturalizmus és az evolúció igazságával szemben. Szalai részletesen bemutatja Plantinga érvét, a különféle válaszstratégiákat, és maga is megfogalmaz egy „kartezianus” ellenérvet Plantinga tézisével szemben.

Huoranszki Ferenc és Sivadó Ákos a természeti fajtákkal kapcsolatos nehézségeket tárgyalja. Huoranszki a természeti fajtáknak azt a Kripke és Putnam nevéhez köthető esszencialista értelmezését bírálja, amely a hetvenes évek elején forradalminak számított az analitikus filozófia korábban általános anti-esszencialista beállítottságához képest. E felfogás szerint az egyedeket esszenciális tulajdonságaik (a kémiai vagy fizikai fajtáknál a mikrofizikai tulajdonságok, a biológiai fajtáknál a genetikai kód) alapján sorolhatjuk természeti fajtákba. Huoranszki elfogadja a naturalista tézist, miszerint vannak természeti fajták, és azt is, hogy ezeket az esszenciális tulajdonságok fogalmával értelmezzük. Viszont azt állítja: a természeti fajták fogalma, Kripke és Putnam javaslatával szemben, nem érthető meg a lehetséges világokra hivatkozó elképzelések segítségével. Az, hogy egy tulajdonság esszenciális egy fajta vonatkozásában, nem azt jelenti, hogy a fajtába tartozó egyedek az illető tulajdonsággal minden lehetséges világban rendelkeznek, hanem azt, hogy az illető tulajdonsággal nem-véletlenszerűen rendelkeznek (ami azonban megengedi, hogy egy a fajtába tartozó egyed – véletlenül – ne rendelkezzen az esszenciális tulajdonsággal).

Sivadó Ákos a biológiai fajták értelmezéseit vizsgálva empirikus oldalról is alátámasztja a Kripke–Putnam-féle esszencialista értelmezés nehézségeit. E felfogás alapján ugyanis olyan sok biológiai faj létezne, hogy az ellehetetlenítené a fajok mint tudományos kategóriák használhatóságát. Sivadó megmutatja, hogy nincs konszenzus a faj értelmezése tekintetében, valamint olyan pluralista elképzeléseket is vizsgál, amelyek szerint nem lehetséges (de nem is szükséges)

egyfajta fajfogalommal leírni a biológiai fajtákat, továbbá amellet is hoz érveket, hogy a társadalmi fajták redukálhatatlanok a biológiai fajtákra.

A természeti fajták értelmezése egyike ama számos kapcsolódási pontnak, amelyek a tanulmányokat összekötik a *Fórum* rovat anyagával. Az itt közreadott válogatás<sup>2</sup> Laki János Thomas Kuhnról szóló monográfiáját (*A tudomány természete*. Budapest, Gondolat. 2006) vitatja meg Forrai Gábor, Pléh Csaba, Schwendner Tibor és Zemlén Gábor recenzióin, illetve a szerző részletes válaszain keresztül. Laki a Kuhnról kialakult kép radikális átértékelését adja. A bevett nézet szerint Kuhn a paradigmaváltás leírása, a megfigyelési és az elméleti dichotómia tagadása, a megfigyelési terminusok elméletterheltségének tézise, vagy az összemérhetetlenséggel kapcsolatos kijelentései révén egy radikális ontológiai relativizmus/antirealizmus képviselője. Ezzel szemben Laki olvasatában Kuhn szerint a tudomány fejlődése racionális, ha nem is a klasszikus logikai pozitivistá, és nem is a popperi értelemben; az elméletváltás nem pillanatszerű, irracionális Gestalt-switch. Az „új Kuhn” továbbá nem antirealista, hiszen „a valósághoz [...] csak a benyomásokból a mindenkori fogalom- és műszerkészlet által létrehozott *episztemikus világ* formájában férünk hozzá, [ugyanakkor] a valóság metafizikai értelemben független a rá vonatkozó vélekedésektől”.

Kuhn (Laki értelmezésében) úgy véli – s itt ismét a Huoranszki és Sivadó által tárgyalt problémákhoz kanyarodunk vissza –, hogy a természeti fajták neveit nem lehet a Kripke–Putnam-féle merev jelölő elmélettel értelmezni, mivel az „azonos természetű” predikátumnak nincs paradigmafüggetlen értelme. Ebből persze rögtön adódnak következmények a természeti fajták értelmezésére is, bár kérdés, hogy pontosan micsodák. Annyi biztos, hogy bár Kuhn Laki értelmezése szerint nem radikális ontológiai relativista vagy antirealista, ha azonban a naturalizmus lényegi részének tekintjük azt a feltevést, hogy léteznek természeti fajták, akkor Kuhnt továbbra sem tekinthetjük naturalistának.

A lapszámot két recenzió zárja Márton Miklós és Klein Dávid tollából, amelyek az utóbbi időben megjelent két fontos metafizikai munkát, Ujvári Márta és Tózsér János írásait mutatják be. Ujvári a nyugati filozófia olyan „örök” problémáit vizsgálja behatóan, mint a szubsztanciák és a változás értelmezése, Tózsér a mai analitikus metafizika legfontosabb témáiba vezet be: e könyvek, egyéb érdemeik mellett, további segítséget nyújthatnak a naturalizmus körüli viták megértéséhez is.

*Ambrus Gergely*

<sup>2</sup> Lásd ezzel kapcsolatban a *Fórum* rovat bevezető sorait.

## A fenomenológia módszertani transzcendentizmusa

Husserl az *Irányadó eszmék egy tiszta fenomenológiához és fenomenológiai filozófiához* első kötetében úgy foglalta össze a transzcendentálfenomenológiai redukció végső eredményét, hogy a fenomenológiában a *realitás* és a *világ* „címszavak bizonyos érvényes *értelmeegységek*”<sup>1</sup> megjelölésére szolgálnak. Ezzel világossá tette, hogy a fenomenológia valóságon és világon nem a tárgyak vagy tények egyszer s mindenkorra adott összességét érti, hanem olyan értelemképződmények együttesét, amelyek szakadatlanul átalakulóban vannak. Ezt az értelemállományt mindazonáltal az *intencionális tudat értelemadó tevékenységére* vezette vissza.<sup>2</sup> E tudatot eközben leválasztotta a dolgok világáról, úgy állítva szembe vele, mint az abszolút lét egy önmagába zárt régióját.

A transzcendentális redukciónak ezt a felfogását alapul véve az értelmezők az *Eszmék* első kötetének megjelenése (1913) óta nem ritkán a husserli fenomenológia *szubjektivista és idealista fordulatáról* szoktak beszélni. Heidegger például – hogy csupán őrá hivatkozzunk – egy előadásának szövegében, amelynek első felét a Husserlrel való számvetésnek szentelte, az *Eszmék* első kötetében „a szubjektivitásnak minden objektivitással szembeni elsőbbségét” vélte felfedezni, és rámutatott „arra a pontra, amelyen az idealizmus, az idealista álláspont, közelebbről a neokantianizmus értelmében vett idealizmus betör a fenomenológiába”.<sup>3</sup>

Ma már nem fér hozzá kétség, hogy az *Eszmék* első kötetében semmiképpen sem valamiféle hirtelen, előre nem látható és csupán átmeneti jelentőségű fordulat játszódik le. A *Husserliana*-sorozat XXXVI. kötetéből világosan kiderül, hogy Husserl nem az *Eszmék* megírásakor, 1912-ben, hanem már jóval korábban, nagyjából 1908-tól kezdve elkötelezte magát a transzcendentális idealizmus mellett. Másrésztől Husserlnek az *Eszmék* első kötetéhez írott késői, 1930-ból származó utószavában a következő áll: „Nem szabad itt [...] elmulasztanom tételesen is

<sup>1</sup> Husserl 1976. 120.

<sup>2</sup> Uo.

<sup>3</sup> Heidegger 1979. 145.

kijelenteni, hogy a transzcendentális idealizmus tekintetében nincs semmi, amit vissza kellene vonnom [...]”<sup>4</sup> Az ez idő tájt papírra vetett *Kartezianus elmélkedések* álláspontja ezzel a kijelentéssel maradéktalanul összhangban van.

A *Husserliana*-sorozat két újabban közzétett kötetéből azonban egyszersmind az is világossá válik, hogy Husserl transzcendentális idealizmuson nem mindig ugyanazt értette, mint az *Eszmék* első kötetében, és nem is mindig úgy érvelt a transzcendentális idealizmus mellett, mint ebben a műben. Egy – a következő vizsgálódások szempontjából irányadó – írásában Rudolf Bernet kiemeli, hogy „Husserl érdeklődése” *A VI. Logikai vizsgálódás átdolgozásához írt vázlatokban*,<sup>5</sup> amelyek 1913-ból származnak, jelentősen elmozdul az *Eszmék* egy évvel korábban írt első kötetéhez képest, és ez az elmozdulás a *VI. Logikai vizsgálódás* módosított változatában Husserl számára lehetővé teszi, hogy „újra elgondolkozzék a fenomenológiai idealizmus értelmén”,<sup>6</sup> és ezzel olyan elmélkedésbe kezdjen, amely nem csupán „pontosabb”, mint az *Eszmék* első kötetének megfelelő gondolatmenete, hanem egyszersmind „kevésbé problematikus” is.<sup>7</sup> Azok az 1914 és 1918 közötti időszakból származó előadászövegek és munkajegyzetek, amelyek a *Husserliana*-sorozat XXXVI. kötetében jelentek meg, megerősítik ezt a benyomást, még ha olykor egymás mellett találunk is bennük olyan érveket, amelyek az *Eszmék* első kötetének gondolatmenetére emlékeztetnek, és olyanokat, amelyek ettől a gondolatmenettől gyökeresen eltérnek. Fontos szerepet játszik ezekben a szövegekben a transzcendentális idealizmusnak az a „bizonyítása”, amelyet Husserl első ízben 1908 táján dolgoz ki, és amelynek azután *A VI. Logikai vizsgálódás átdolgozásához írt vázlatokban* új megfogalmazást ad. Különös figyelemre tarthat számot a *Husserliana* XXXVI. kötetének ötös, hatos, hetes és nyolcas számú szövege – és köztük is mindenekelőtt a hetes számú szöveg –, mert ezekben Husserl az említett bizonyításból olyan következtetéseket von le, amelyek felvetik a kérdést: vajon egyáltalán helytálló-e a fenomenológia módszertani transzcendentális idealizmusát *idealizmusként* megjelölni.

A következő vizsgálódásokban arra teszek kísérletet, hogy ezt a kérdést *mint* kérdést exponáljam. Három lépésben haladok előre e cél felé: először is megkísérlem elhatárolni Husserlnek a transzcendentális idealizmus mellett a XXXVI. kötet írásaiban ismételen felhozott érvét az *Eszmék* első kötetében található gondolatmenettől; másodsor az említett érv szerkezetét igyekszem feltárni; végül pedig azokat a megfontolásokat veszem sorra, amelyek kétségeket támaszthatnak bennünk afelől, hogy a fenomenológia módszertani transzcendentális idealizmus vajon szükségszerűen összekapcsolódik-e a transzcendentális idealizmus álláspontjával.

<sup>4</sup> Husserl 1971. 150. sk.

<sup>5</sup> Husserl 2002.

<sup>6</sup> Bernet 2004. V. fejezet, 143–168., különösen 147.

<sup>7</sup> Uo. 146.

## I. KÉT KÜLÖNBÖZŐ ÉRV A TRANSCZENDENTÁLIS IDEALIZMUS MELLETT

Az *Eszmék* első kötetében az úgynevezett „fundamentális fenomenológiai vizsgálódással” Husserlnek az a fő célja, hogy különbséget tegyen a „*lét mint élmény*” és a „*lét mint dolog*” – vagyis az élmény és a dolog értelmében vett lét – között.<sup>8</sup> A Husserl szeme előtt lebegő különbség abból adódik, hogy a (téridőbeli) dolog „transzcendens észlelés” tárgya, az élmény viszont „immanens észlelésben” válik megragadhatóvá.<sup>9</sup> Ez utóbbi észlelést azért nevezzük „immanensnek”, mert olyan élményt tesz intencionálisan (*intentional*) tárgyává, amellyel egy és ugyanazon élményfolyamban valóban (*reell*) is összetartozik. Az észlelés tárgya itt következképpen *valós része* az észlelés aktusának: ténylegesen is *benne rejlik* ebben az aktusban. Éppen ezt a „benne rejlik” kifejezéssel jellemzett kapcsolatot hivatott megjelölni az „immanencia” terminus.<sup>10</sup>

Kétségtelen ugyan, hogy a (téridőbeli) dologra vonatkozó és azt sikeresen megragadó észlelésben ez a dolog úgyszintén a *maga közvetlen jelenvalóságában*, sőt *testi valójában* válik adottá.<sup>11</sup> Sohasem észleljük azonban a (téridőbeli) dolgot *a maga teljességében*; egy ilyen dolog mindig csak aspektusaiban vagy, ahogyan Husserl mondja, tárgyi aspektusaiban avagy „árnyalataiban” (*Abschattungen*jaiban) válik észlelhetővé.<sup>12</sup> Ez a megfigyelés rávilágít arra a perdöntő különbségre, amely „a lét mint dolog” és „a lét mint élmény” között megvonható. Husserl így fogalmazza meg ezt a különbséget: „*Egy élmény nem bomlik tárgyi árnyalatokra [schattet sich nicht ab].*”<sup>13</sup>

Eddigi megfontolásaink azt a benyomást kelthetik, mintha transzcendens észlelésen olyan észlelést kellene értenünk, amely tárgyait nem a maguk teljességében ragadja meg, és amelyet éppen ezért *inadekvát* kell neveznünk. Csakhogy ez a benyomás csalóka. Husserl ugyanis hangsúlyozza: „Az élményt sem észleljük sohasem a maga teljességében; teljes egységében adekvát módon az élmény sem ragadható meg.”<sup>14</sup> Transzcendens és immanens észlelés különbsége tehát nem esik egybe inadekvát és adekvát észlelés különbségével.

Ezért további vizsgálódásokra van szükségünk ahhoz, hogy kideríthessük, miben is áll ez a különbség. Eddigi megfontolásaink fő eredménye az a megfigyelés, hogy a téridőbeli dolog – az élménnyel ellentétben – alá van vetve a perspektivikus rövidülés törvényének, és ezért mindig csak tárgyi aspektusai-

<sup>8</sup> Husserl 1976. 87.

<sup>9</sup> Uo.

<sup>10</sup> Uo. 78. sk.

<sup>11</sup> Uo. 81.

<sup>12</sup> Uo. 84. sk.

<sup>13</sup> Uo. 88.

<sup>14</sup> Uo. 93.

ban avagy árnyalataiban válik észlelhetővé. Ki kell azonban még bontanunk e tény további következményeit.

Egyenes következménye e ténynek, hogy a dologészlelés szükségszerűen *prezumptív jellegű*. Ezen azt kell értenünk, hogy ámbár egy dolog észlelése egyértelműen az észlelt dolog tényleges létének feltevése mellett szól, sohasem zárható ki egészen, hogy a dolog létezését az újabb észlelések mégiscsak megkérdőjelezzék majd. Ebből a meglátásból kiindulva fontos megállapításra jutunk: „Az adottságból sohasem következik szükségszerűen a dologi lét, hanem ez a fajta lét bizonyos módon mindig *esetleges* marad.”<sup>15</sup> A transzcendens észlelés szükségszerűen léttételező jellegű ugyan, ámde az általa tételezett lét mindenkor *esetleges*. Más a helyzet az immanens észleléssel: ez a fajta észlelés kezeskedik intencionális tárgyának valós létéért, úgyszólván jótáll érte, azaz: *szükségszerűvé*<sup>16</sup> teszi e tárgy valós létét.

Így a transzcendens és immanens észlelés között az *Eszmék* első kötetében megvont különbség *esetleges és szükségszerű lét* szembeállításához vezet. Innen már csak egy további lépésre van szükség ahhoz a következtetéshez, hogy a dologi lét mindig *relatív*, mert rá van utalva a vele intencionális kapcsolatba lépő élményre, és hogy az élmény léte ezzel ellentétben *abszolút* lét. E megfontolásból kiindulva jut el Husserl ahhoz a felfogáshoz, amely szerint a tiszta tudat „*önmagáértvalóan zárt létösszefüggés*”.<sup>17</sup> Ezzel az *Eszmék* első kötetében a tiszta tudat „régioja” teljes mértékben elkülönül a transzcendens (téridőbeli) dolgok régiójától. A mű híres – vagy inkább hírhedt – 49. §-a szerint az abszolút létnek ezt a régióját a világ megsemmisülése sem érintheti.<sup>18</sup> Husserl ebből kiindulva végül is a következő tételhez érkezik el: „Az immanens lét tehát kétségtelenül abban az értelemben abszolút lét, hogy elvi okoknál fogva *nulla »re« indiget ad existendum*.”<sup>19</sup> Ehhez még hozzáfűzi, hogy „a transzcendens »res« világa teljességgel a tudatra, mégpedig nem a logikailag elgondolt, hanem az aktuális tudatra van ráutalva”.<sup>20</sup>

Az *Eszmék* első kötetében ez a két tétel rögzíti a transzcendentális idealizmus jelentését. Sokan figyeltek már arra, hogy az e tételekhez vezető gondolatmenet karteziánus megalapozottságú. Az okfejtés azonban korántsem azért kétes színezetű, mert karteziánus alapokra támaszkodik, hanem inkább azért, mert nem vet kellőképpen számot a karteziánus én tényszerű adottságával. Husserl az *Eszmék* első kötetében világosan felismeri, hogy a *cogitót* csupán „egy faktum szükségszerűsége” illeti meg.<sup>21</sup> Olyan szükségszerűséget ért ezen, amely alap-

<sup>15</sup> Uo. 97.

<sup>16</sup> Uo. 98.

<sup>17</sup> Uo. 105.

<sup>18</sup> Uo. 104.

<sup>19</sup> Uo. [Nincs ráutalva semmilyen „dologra” ahhoz, hogy létezzon.]

<sup>20</sup> Uo.

<sup>21</sup> Uo. 98.

vetően különbözik minden „tiszta lényegi szükségszerűségtől”, a „lényegtörvények [mindenfajta] konkrétizációjától”.<sup>22</sup> De vajon kellőképpen teherbíró-e ez a merőben tényszerű szükségszerűség ahhoz, hogy hordozhassa esetleges és szükségszerű létezés szembeállításának ontológiai terhét? Az *Fszmék* első kötetének megírásakor Husserl nyilvánvalóan így véli. Véleménye azonban később megváltozik: arra a belátásra jut, hogy az én öneszmélése – a *cogito* – maga is „az ösesetlegesség magvát” hordozza magában.<sup>23</sup>

A *Husserliana*-sorozat XXXVI. kötetének írásaiban a transzcendentális idealizmus bizonyítására több helyütt egészen más felépítésű érv szolgál, mint az *Fszmék* első kötetében. Ez az érv már csak azért is felkeltheti az érdeklődésünket, mert sokkal inkább számot vet a *cogito* tényszerű szükségszerűségével, mint az *Fszmék* első kötetében található gondolatmenet. A bizonyítás ezúttal nem arra a megfontolásra épül, hogy a dologi lét egyoldalú módon rá van utalva az élmény léteire. Husserl most egészen másfajta szembeállításra támaszkodik: *reális* és *ideális lehetőség* között tesz különbséget. Ez a megkülönböztetés inkább Leibnizre, mint Descartes-ra emlékeztet. Az érv *reális létezés* és *aktuális tudat* specifikus kapcsolatában csúcsosodik ki, amely a megfontolások szerint különbözik a lehetséges tárgyak és a nekik megfelelő tudatformák általános korrelációjától. Amint a *Husserliana*-sorozat XXXVI. kötetének szerkesztői állítják, itt a transzcendentális idealizmusnak valamiféle „speciális” tézisééről van szó, amely nem tárgy és tudat egyáltalában vett korrelációját mondja ki, hanem azt a sajátos függőséget hangsúlyozza, amely a *reális* létezés és az *aktuális*, azaz *faktikusan létező* tudat között áll fenn. Azt az érvet, amely e tézis igazolására szolgál, a továbbiakban „speciális érvnek” fogom nevezni, hogy terminológiailag is megkülönböztethessem attól az okfejtéstől, amelyet az *Fszmék* első kötetében találunk. Az utóbbi okfejtés felépítésére az előzőekben már rávilágítottunk. Gondolattmenetünk következő lépése megkívánja, hogy hasonló módon feltárjuk a „speciális” érv szerkezetét is.

## II. A „SPECIÁLIS” ÉRV SZERKEZETE

Husserl abból indul ki, hogy fenomenológiailag tekintve a lét *igazolhatóság*ot jelent. Az igazolhatóság szerinte összefüggésben van a szemlélettel. A *Husserliana*-sorozat XXXVI. kötetének ötös számú szövegében erről a következőket olvashatjuk: „[...] ami van, az szükségképpen adottsághoz juttatható; minden lehetséges tárgy [...] korrelátuma egy lehetséges szemlélet [...]”.<sup>24</sup> Az „igazolhatóság elve” Husserl gondolatmenetének egy tovább már nem megalapozott és

<sup>22</sup> U.o.

<sup>23</sup> Husserl 1973. 386. (22. számú szöveg).

<sup>24</sup> Husserl 2003. 73.

nem megalapozható premisszája. De talán nem is igényel további megalapozást ez az elv, hiszen közvetlenül adódik a fenomenológia legalapvetőbb meggyőződéséből, amely így fogalmazható meg: egy szemléleti tartalom nélküli, vagyis üres intenciónak lehet ugyan értelme és jelentése, ámde egy ilyen intenció önmagában nem képes intencionális tárgyának létét biztosítani; csakis az ezen üres intenciót betöltő „szemlélet” – a kanti értelemben vett „korrespondáló szemlélet” – kezeshet a tárgy létéért.

A *Husserliana*-sorozat XXXVI. kötetének írásaiiban a transzcendentális idealizmus bizonyítása azáltal kezd határozottabb alakot öltetni, hogy Husserl nem elégszik meg lehetséges tárgy és lehetséges szemlélet korrelációjának megállapításával, hanem az igazolás reális és ideális lehetősége közti különbségre is kitér. A téridőbeli dolgok reális létezésére alkalmazva az igazolhatóság elve korántsem olyan magától értetődő, mint általános alakjában. Ez a különbség teszi szükségessé a transzcendentális idealizmus „speciális” tézisének kimondását. Hogy ezt nyilvánvalóvá tehesük, idézzük föl, miben is különbözik egymástól az igazolás reális és ideális lehetősége.

Ideális tárgyak, így például számok vagy geometriai alakzatok esetében az igazolásnak csupán az ideális lehetősége jön szóba. Ez azonban egyszersmind már kezeshet is az ilyesfajta tárgyak – igaz, csupán ideális – létéért. Csakhogy: „[...] a számok, mennyiségek stb. matematikai létezése az evidens igazolás ideális lehetőségének feltételeként nem igényli okvetlenül egy olyan tudat valóságos létezését, amely közvetlenül vagy közvetve matematikai tárgyakra irányulna, vagy amelynek ilyen tárgyakra kellene irányulnia.”<sup>25</sup>

Hasonló megállapítás érvényes a képzelet tárgyaira. A *Husserliana*-sorozat XXXVI. kötetének ötös számú szövegében Husserl a kentaurt említi példaként olyan tárgyakra, amelyeket szabadon el tudunk képzelni, a hatos számú szövegben pedig hasonló értelemben ejt szót a sellőről.<sup>26</sup> Kétségtelen, hogy az ilyesfajta fantáziaképződmények lényegileg különböznek az ideális tárgyaktól. A számokkal és a geometriai alakokkal ellentétben ezek nem eidetikusan általános, hanem individuális tárgyak, amelyek a képzeletben éppúgy alá vannak vetve a perspektivikus rövidülés törvényének, akárcsak a téridőbeli dolgok a valóságban. Ebből következőleg a kentaur ideális lehetőségéről csak úgy győződhetem meg, hogy „a fantáziában a fiktív alakot elém táró szemléleti aktusokat egymással egybehangzó módon végrehajtom; a fantáziában közelebb megyek ehhez az alakhoz, minden oldaláról megtekintem, megtapogatom, és létezését minden módon vizsgálatnak vetem alá stb.”<sup>27</sup> Az ilyesfajta képzeletbeli vizsgálatra csak azért van mód, mert a fantáziában – amint ez a *Husserliana*-sorozat XXIII. kötetének bizonyos fejtegetéseiből egyértelműen kiderül – *fantázia-énem* egy-

<sup>25</sup> Uo. 74.

<sup>26</sup> Uo. 113.

<sup>27</sup> Uo. 75.



szersmind *fantáziatettel* is rendelkezik. A képzeletbeli alakok ideális lehetősége mindazonáltal nem feltételez aktuális tudatot és valóságosan létező ént. Ennek az oka, hogy a fantázia-ént nem azonos a valóságos énnel. A sellő példáján ez különösen világosan meg is mutatkozik: velem, „az itt és most aktuálisan eleven énnel, aki ebben a szobában vagyok – mondja Husserl –, az elképzelt sellő például nem valóságosan áll szemben; kedves arcát, majd pedig büszkén a hátát nem felém, faktikus énem felé fordítja stb. Amik felém, faktikus énem felé fordulnak, és erről vagy arról a tájékról körülöttem rám tekintenek, valamely oldalukkal elélem tárulnak, azok valóságos környezetem valóságos dolgai. De vajon nem állíthatjuk-e: »azáltal, hogy a képzeleti térbe, az elképzelt tájba, amelyben az elképzelt sellő megjelenik előttem, magamat is beleképezem, magam is ott vagyok a közelében«? Lehetséges. Csakhogy ezáltal a képzeletben feladom aktuális létezésemet, és megszűnök *valóságos énnel* lenni.”<sup>28</sup> Husserl ebben a gondolatmenetben nyilvánvalóan arra a megfigyelésére támaszkodik, amely észlelet és képzelet egymást kizáró ellentétére (*Widerstreitjére*) vonatkozik, és amelyet már az 1904–05-ös téli szemeszterből származó nevezetes előadásában is megfogalmazott. Észlelet és képzelet egymást kizáró ellentétéből adódik az a következtetés, hogy a valóságos tárgyakhoz hasonlóan perspektivikus rövidülést mutató fantáziatárgyak testileg is meghatározott ént feltételeznek ugyan, ámde mégiscsak olyan ént, amely nem esik egybe a valóságos énnel, hanem maga is a képzelethez tartozik.

A képzelet közelebbi elemzése azért tartozik hozzá lényegileg a transzcendentális idealizmus husserli bizonyításához, mert nyilvánvalóvá teszi, hogy az ideális lehetőségek birodalmában az egyöntetűségnek teljességgel egyenértékű rendszerei (*Einstimmigkeitssysteme*) állhatnak fenn, anélkül, hogy az egyik rendszer kitüntetettséget élvezne a másikkal szemben. Minden ideális lehetőség *egyformán lehetséges*, még ha egyébként az egyik kizárja is a másikat;<sup>29</sup> ez utóbbi körülményből csupán az következik, hogy nem állhatnak fenn *együttesen*, ettől azonban még egyik sem élvez előnyt a másikkal szemben. Másként áll a dolog a reális lehetőséggel: ez kizárja az egyöntetűségnek minden olyan rendszerét, amellyel ő maga nem fér össze. Azt is mondhatnánk, hogy amikor egy ideális lehetőség valóra válik – *realizálódik* –, nyomban minden más ideális lehetőséget félreszorít. De mi az, ami meghatározza, hogy az ideális lehetőségek közül melyik valósulhat meg, és melyik nem?

Ez a kérdés átvezet minket a fantáziatárgyaktól a reálisan létező dolgokhoz. A feltett kérdésre ugyanis Husserl szerint végső soron csak egy választ adhatunk: a dolog a maga teljességében és mindenoldalú meghatározottságában egyszerre mind kezünbe adja azt a vezérfonalat is, amelynek segítségével az egyöntetűség ideálisan egyformán lehetséges rendszerei közül ki tudjuk választani azt az egy

<sup>28</sup> Uo. 113.

<sup>29</sup> Uo. 75.

rendszer, amely a tárgy reális létezése esetén egyedül tekinthető valóságosnak. Nem nehéz szemléletesen megvilágítani e válasz értelmét: ha a kentaur nem pusztán fantáziaképződmény volna, hanem reálisan létező lény, akkor olyan kérdések, mint például hogy milyen színárnyalatú a kentaur szőre, vagy hogy egy-egy kétszemű lény-e a kentaur, egyértelműen eldönthetők volnának.

Ámde ezen a ponton felmerül egy ellenvetés: amint arról már szó volt, egyetlen reálisan létező dolog sem mutatkozik meg számunkra a maga teljességében és mindenoldalú meghatározottságában, hanem csupán tárgyi aspektusaiban avagy árnyalataiban válik észlelhetővé. Husserl szerint azonban ez a tény nem változtat azon, hogy a dolognak *valamiképpen* mégiscsak megvan a maga teljes adottsága és mindenoldalú meghatározottsága. A kérdés csupán az: *miképpen*. Husserl egyértelmű választ ad erre a kérdésre: szerinte a dolog *adekvát* adottsága *kanti értelemben vett eszme*, amelynek a tapasztalatszerzés során regulatív, szabályozó szerepe van. Ez az eszme alakítja ki bennünk azt a meggyőződést, hogy a dolog önmagában valóan mindenoldalúan meghatározott.

Amint ismeretes, Husserl a magábanvalósága szerint tekintett dolgot már az *Eszmék* első kötetében is kanti értelemben vett eszmeként fogja fel. Rudolf Bernet joggal utal arra, hogy egy kanti értelemben vett eszme nem tekinthető „eidosz”-nak, általános lényegnek, hanem csupán egy egyedi dolog eszméjeként gondolható el.<sup>30</sup> Nem csupán *A VI. Logikai vizsgálódás átdolgozásához írt vázlatok*, hanem a *Husserliana*-sorozat XXXVI. kötetében összegyűjtött szövegek is egyértelműen azt mutatják, hogy különbséget kell tennünk eidosz és kanti értelemben vett eszme között. Erről Husserl egy helyütt a következőket írja: „[...] *a dolog léte* az aktuális tudat számára mindenkor egy *eszme*, ámde nem egy tisztán ideális lét értelmében vett eszme, mint egy szám, egy *species*, hanem *kanti értelemben vett* (többdimenziós) eszme [...]”<sup>31</sup> Ezekből a megállapításokból olyan elképzelés rajzolódik ki, amely szerint valahányszor megkülönböztetjük az önmagában valóan tekintett dolgot a rá vonatkozó észleletektől és gondolatoktól, valójában a hozzá tartozó jelenségek avagy tárgyi árnyalatok mindenoldalúan végtelen kontinuumát tartjuk szem előtt. Husserl úgy látja, hogy a dolog a maga egyszeri és megismételhetetlen mivoltában nem más, mint e jelenségkontinuum a maga végtelen teljességében.

Abban a gondolatban, hogy a reális dolog a maga egyedi és megismételhetetlen létezésében nem egyéb, mint kanti értelemben vett eszme, a transzcendentális idealizmus mellett szóló „speciális” érv kulcsgondolatát fedezhetjük fel. Ebből a gondolatból már következik, hogy egy reálisan létező dolog nem pusztán egy ideálisan lehetséges, hanem egy aktuális, valóságos, faktikusan létező tudatot feltételez. Ahogyan Husserl maga fogalmaz: „*egy dolog létezésének eszméje*

<sup>30</sup> Bernet 2004. 161.

<sup>31</sup> Husserl 2003. 77.

előírja, hogy létezzék egy valóságosan tapasztaló tudat is [...]”.<sup>32</sup> Ez az összefüggés abból adódik, hogy egy reális dolog egyedi létezésének eszméje végső soron nem egyéb, mint az egyöntetűség egy mindenoldalúan végtelen rendszerének eszméje; egy ilyen rendszer azonban csak akkor tekinthető egyértelműen meghatározottnak, ha a valóságos tapasztalat folyamata kitünteti a merőben ideális lehetőségek közül. Nyilvánvaló ugyanis, hogy csupán a valóságos, faktikusan végbemenő tapasztalatszerzés nézőpontjából tekinthetők kitüntettnek azok a „reális motivációs lehetőségek”,<sup>33</sup> amelyekből az egyöntetűségnek egy mindenoldalúan végtelen rendszere felépül.

Ezzel világossá vált előttünk a transzcendentális idealizmus mellett szóló „speciális” érv szerkezete. A bizonyítandó tétel azt mondja ki, hogy a dolog reális létezése nem független egy aktuális, valóságos, faktikusan létező tudattól. Az egész gondolatmenet arra a feltevésre támaszkodik, amelyet „az igazolhatóság elveként” rögzítettünk. De a bizonyítás középponti eleme – avagy „idegyszála” – mégiscsak az a meglátás, hogy egy dolog reális létezése a maga egyszeri és megismételhetetlen egyediségében nem egyéb, mint kanti értelemben vett eszme, amely – mint „regulatív eszme” – csak a faktikusan végbemenő tapasztalat végtelen előrehaladásában válik valóra. Ezzel a „speciális” érv a valóságosan tapasztaló tudatot olyan faktumként – avagy olyan *őstényként* – mutatja fel, amely minden realitás feltétele.

### III. MÓDSZERTANI TRANSCZENDENTALIZMUS ÉS TRANSCZEDENTÁLIS IDEALIZMUS

A transzcendentális idealizmus imént összefoglalt bizonyítása döntően újat hoz az *Eszmék* első kötetének okfejtéséhez képest. Mindazonáltal az 1914 és 1918 között keletkezett előadásszövegekben és munkajegyzetekben a régi gondolatmenet nyomait is megtaláljuk. Husserl új érvvel próbálkozik ugyan, ámde eközben újra meg újra visszanyúl az egyszer már a nyilvánosság előtt is vállalt és olykor továbbra is biztosnak érzett régi felismerésekhez. Így például a hatos számú szövegben ezt olvashatjuk: „A tudat léte alapvetően megkérdőjelezhetetlen. Szükségszerű lét. A világ léte esetleges, a dolog úgy áll, hogy világnak egyáltalán nem kell lennie.”<sup>34</sup> Könnyű felismerni ezekben a lapidáris kijelentésekben az *Eszmék* első kötetéből ismert okfejtés fő állításait.

Ezzel egyidejűleg azonban a *Husserliana*-sorozat XXXVI. kötetének írásaiban merőben új szempontok is felmerülnek. A transzcendentális idealizmus mellett szóló „speciális” érvből olyan következmények is adódnak, amelyek egyrészt

<sup>32</sup> Uo.

<sup>33</sup> Uo.

<sup>34</sup> Uo. 111.; vö. 124.: „A világnak nem szükséges lennie”.

a testiségre, másrészt pedig az interszubbjektivitásra vonatkoznak. Egyrésztől Husserl már a hatos számú szövegben is olyan megfogalmazást választ a transzcendentális idealizmus kifejezésére, amely inkarnált, testiséggel is rendelkező tudatot implikál: „Minden dolog *a priori* egy *aktuális* én környezetében található.”<sup>35</sup> A „környezet” szó azt sejteti, hogy a téridőbeli dolgok olyan ént feltételeznek, amely környező világának középpontjában (orientációs zérópontjában) helyezkedik el, és ennél fogva szükségszerűen testi meghatározottsággal rendelkezik. Másrésztől a világnak a végtelenbe nyúló tapasztalati megismerése, amelyre a kanti értelemben vett eszme már fogalmánál fogva is utal, csakis egy interszubbjektív megismerő közösség műve lehet. Mindkét következmény megfogalmazódik már *A VI. Logikai vizsgálódás átdolgozásához írt vázlatokban* is, hogy azután újabb kutatások tárgyává válják a *Husserliana*-sorozat XXXVI. kötetének hetes számú szövegében, amely 1914-ből vagy 1915-ből származik.<sup>36</sup>

Ebben a hosszabb terjedelmű feljegyzésben Husserl nem elégszik meg annyival, hogy érvényre juttassa testiség és interszubbjektivitás két új szempontját. Jóval tovább megy ennél: merőben új megvilágításba helyezi a transzcendentálfenomenológiai idealizmus egész értelmét. „A megismerés szubbjektumának, amelynek [a transzcendentális idealizmus mellett szóló »speciális« érv szerint] valóságosan létezőnek kell lennie, ugyanakkor nem szükséges aktuálisan is megismerőnek lennie [...]”<sup>37</sup> – jelenti ki; ez az állítás azonban még nem is annyira meglepő. Nagy horderejű meglátás fejeződik ki viszont abban, amit Husserl ehhez az állításhoz hozzáfűz: „Arra sincs szükség azonban, hogy egy ilyenfajta szubbjektum a világ meglétének egész végtelen időtartama alatt létezzék.”<sup>38</sup> Husserl ezt a gondolatot a következőképpen fejti ki: „Egy pusztán materiális világ mint a világ alsóbb rétege és mint időtartamának kezdeti szakasza eleget tesz a megismerhetőség feltételeinek, ha létezik egy szubbjektivitás, amely tapasztalat és gondolkodás révén konstituálja azt a világot, amely az ő jelenlegi környező világa, és emellett az ésszerűség követelményeinek megfelelően egyszersmind visszamenőleg is konstituálni képes az elmúlt világszakaszokat és ezek között esetleg a pusztán materiális természetnek egy szakaszát is.”<sup>39</sup>

Azt gondolhatnánk, hogy itt talán nem a transzcendentális szubbjektivitásról van szó, hanem csupán az immár emberként appercipiált szubbjektumokról. A hetes számú szöveg táplálni látszik ezt a gyanút, hiszen a vizsgálódások fő témája itt – amint Husserl kiemeli – a „testi, emberi szubbjektivitás”.<sup>40</sup> Ezzel a tárgymegjelöléssel összhangban áll, hogy Husserl egy helyütt „a természet em-

<sup>35</sup> Uo. 114.

<sup>36</sup> Uo. 132–140.

<sup>37</sup> Uo. 140.

<sup>38</sup> Uo.

<sup>39</sup> Uo. 141.

<sup>40</sup> Uo. 132.

ber előtti múltját” említi.<sup>41</sup> A *visszamenőleges konstitúció* gondolata viszont mégiscsak félreérthetetlenül arra utal, hogy az idézett gondolatmenet a transzcendentális szubjektivitásra vonatkozik. Ezen értelmezés mellett szól az a körülmény is, hogy a transzcendentális idealizmus bizonyítása a dolgok reális létezését természetszerűleg nem az emberiként apperceptiált, hanem a transzcendentális szubjektivitástól teszi függővé.

Kétségtelen, hogy a visszamenőleges konstitúció gondolata a legkevésbé sem magától értetődő. A világ a fenomenológia nézőpontjából tekintve szükségszerűen konstitúció eredménye. De ha ez így van, vajon létezhet-e egyáltalán világ akkor, amikor „a konstitúció nem aktuálisan végbemenő konstitúció”?<sup>42</sup> Azokhoz a megfontolásokhoz, amelyekkel Husserl e kérdésre választ keres, tagadhatatlanul tapad némi kísérleti jelleg. Végső soron azonban kikristályosodik belőlük egy olyan válasz, amely egyértelműen a visszamenőleges konstitúció eszméjének megerősítéséül szolgál: „Egy világ olyan szubjektumok nélkül, amelyek e világról tapasztalatot szerezhetnének (más szóval ezt a világot térben, időben és oksági meghatározottságában szemlélhetnék), csak egy ilyen szubjektumokkal rendelkező világ múltjaként gondolható el.”<sup>43</sup> Husserl itt nyilvánvalóan olyan múltra gondol, amelyet a szóban forgó szubjektumok „visszamenőlegesen konstituálnak”.

A visszamenőleges konstitúció gondolatából egy olyan álláspont körvonalai bontakoznak ki, amelyet *módszertani transzcendentalizmusként* jelölhetnénk meg. Ez az álláspont a tudat őstényére alapozódik, és a reális lét – avagy a dologi létezés – kérdését az intencionális konstitúció nézőpontjából veti fel. Ebből következik, hogy aki a módszertani transzcendentalizmus mellett kötelezi el magát, az éppúgy szembehelyezkedik a naiv realizmussal, mint a dogmatikus materializmussal avagy naturalizmussal. Az intencionális konstitúció nézőpontja azonban nem zárja ki, hogy számot vessünk a valóságnak egy „alsóbb rétegével”, amely mint a világlétezés kezdeti szakasza akkor is létezik, *amikor a konstitúció még nem aktuálisan végbemenő konstitúció*, és amely éppen ezért csakis egy *visszamenőleges konstitúció* tárgyaként válhat megragadhatóvá.

Nem könnyű eldönteni, vajon ez a tisztán módszertani transzcendentalizmus, amelyhez Husserl a *Husserliana*-sorozat XXXVI. kötetének hetes számú szövegében elérkezik, egyáltalán értelmezhető-e még (transzcendentális) idealizmusként. Egyrészt ugyanis Husserl továbbra is ragaszkodik ahhoz, hogy a dolgok reális létezése aktuális tudatot feltételez. Ez a transzcendentális idealizmus tézisének fenntartása mellett szól. Másrészt azonban Husserl most arra a belátásra jut, hogy van a világnak egy „alsóbb rétege”, amely akkor is létezik, ha nem tárgya aktuális konstitúciónak. Ezzel olyan realitást tár a szemünk elé, amely

<sup>41</sup> U.o. 143.

<sup>42</sup> U.o. 141.

<sup>43</sup> U.o. 144., 2. megj.

*magában a tudatban bizonyul a tudattól függetlennek.* Mindez arra vall, hogy a tudat őstényének mélyebb elemzése megmutathatja, miképpen utal a tudat önmagán túl olyasvalamire, ami *független* tőle.

Ezt az eredményt úgy is értelmezhetjük, mint az idealizmus és a realizmus, valamint a szubjektivizmus és a naturalizmus egyidejűleg történő meghaladását. Ezeket az álláspontokat nem foglalhatjuk el anélkül, hogy túl ne lépnénk annak körén, ami adott számunkra. Husserlnek, úgy gondolom, igaza van abban, hogy nem tekinthetünk el – és nem is vagyunk képesek eltekinteni – saját tudatunk élményszerű adottságától, amely megadja számunkra a világ egyes szám első személyben megfogalmazódó tapasztalatát, belső nézetét, szubjektív vetületét. Husserl azért utasítja el a realizmust és a naturalizmust, mert olyan álláspontokat lát bennük, amelyek azt várják el tőlünk, hogy tudatos tevékenységet folytassunk, eközben azonban ne vegyünk tudomást saját tudatunk élményszerű adottságáról. Nincsenek kétségei afelől, hogy lehetséges a világlétezésnek egy olyan kezdeti szakasza, amelyben nem létezik tudat, de módszertanilag kizártnak tekinti, hogy a világlétezésnek ezt a kezdeti szakaszát úgy ragadjuk meg, mintha ma sem léteznék tudat. A tudat megkerülhetetlen élményszerű adottságát azonban nem téveszti össze a tudat létszerű szükségyszerűségével. Az a fajta transzcendentális idealizmus, amely mellett elkötelezi magát, nem abszolút idealizmus, vagyis nem olyan felfogás, amely egy szellemi létezőnek saját tudatunk élményszerű adottságától független létezését állítja. Husserl egész életművének legfőbb célját abban látja ugyan, hogy fényt derítsen a szubjektivitás tüneményére, ámde ebből nem következik, hogy a világ szubjektív vetületét tekintően az egyedüli valóságnak, és így a szubjektivizmus csapdájába esnek.

Félreértés volna azt hinni, hogy a módszertani transzcendentalizmus kizárja az objektív leírásra és magyarázatra törekvő tudományt. Késői korszakában Husserl ahhoz a meggyőződéshez érkezik el ugyan, hogy az életvilág előzetes megragadása nélkül a világ objektív megragadására irányuló törekvés értelmét veszíti. A módszertani transzcendentalizmussal eleve összeférhetetlen az a hiedelem, hogy a valóság megértésének kulcsát a világlétezés tudat nélküli szakaszának megragadása vagy a tudat őstényéről tudomást nem vevő magyarázata adhatná kezünkbe. A fenomenológia szerint a valóság elsődlegesen az életvilág valósága, az életvilág pedig nem képzelhető el saját tudatunk élményszerű adottsága nélkül. Mindazonáltal Husserl nem gondolja azt, hogy az objektív tudományok ne *bővíthetnék* és *helyesbíthetnék* azt a képet, amely az egyes szám első személyben megfogalmazódó tapasztalat alapján bennünk az életvilágról kialakul. A fenomenológiának nem kell minden értelemben elutasítania azt az elképzelést, hogy az objektív tudományok – szemben az egyes szám első személyben megfogalmazódó tapasztalattal – a *magában való realitást* ragadják meg. Az a gondolat, hogy az objektív tudományok a magában való realitást *abszolút* értelemben ragadják meg, nem fér össze a módszertani transzcendentalizmussal. Ezzel szemben nagyon is összefér a módszertani transzcendentalizmussal az

a felfogás, amely szerint az objektív tudományok *relatív* értelemben mégiscsak a magában valót ragadják meg, ha ezen azt értjük, hogy képesek *kiegészíteni* és *helyreigazítani* a világnak azt a szubjektív vetületét, amely az egyes szám első személyben megfogalmazódó tapasztalat alapján rajzolódik ki bennünk.

A módszertani transzcendentalizmus elvéből ezenfelül olyan látásmód következik, amely a valóságot *rétegekre* bontja, és a magasabb szinteknek (a tudatéletnek és a történeti valóságnak) az alacsonyabbakból (az élettelen és az élő természetből) történő levezetésére éppúgy kételyekkel tekint, mint arra a fordított irányú elképzelésre, hogy az alacsonyabb rétegek eleve a magasabbak kedvéért jöttek volna létre. Husserl főként az *Ideen II*-ben dolgozta ki a regionális ontológiáknak ezt a rétegzett felfogását. Hasonló álláspontra helyezkedett Nicolai Hartmann is, aki négykötetes ontológiájának harmadik kötetében világosan meghatározta azokat az alapelveket is, amelyek ezt a valóságot rétegekre bontó látásmódot jellemzik. Érdekes itt ezek közül két alapelvet kiemelni. Hartmann szerint egyrészt a magasabb szintek *függőségben* vannak az alacsonyabb szintektől, másrészt azonban az alacsonyabb szintekhez képest *újat* – „kategorialis nóvumot” – hoznak az magukkal. A rétegotológia mellett Husserl és Hartmann szemében főként az szól, hogy általa lehetségessé válik az egyes szintek tudományos és filozófiai megragadása anélkül, hogy az egyes szintek keletkezésének megválaszolhatatlan – és ezért ideologikus spekulációkra csábító – kérdéseibe kellene belemerülnünk.

Az idealizmus és a realizmus, valamint a szubjektivizmus és a naturalizmus közti vita valójában a *világnézetek* között folyó harc része. Politikai elkötelezettségeink és esetleges vallási meggyőződéseink nyilvánvalóan nem függetlenek e harc mindenkori állásától. Egy módszertanilag átgondolt munkafilozófiának azonban semmi dolga a világnézetek között folyó harccal. Nemcsak Husserl, hanem – főleg korai alkotói periódusában – Heidegger is kiáll filozófia és világnézet egyértelmű elválasztása mellett. A fenomenológia módszertani transzcendentalizmusa éppen abban különbözik a hagyományból jól ismert idealizmustól, hogy nem értelmezhető világnézeti állásfoglalásként. Ezért, ha a fenomenológiáról ejtünk szót, talán bölcsőbb volna felhagynunk a „(transzcendentális) idealizmus” terminus használatával. Ha lemondunk erről a megjelölésről, az maradéktalanul összhangban van azzal a meggyőződéssel, hogy egy pluralista módon és demokratikusan berendezett világban a filozófia a világnézeti kérdésekben való állásfoglalást egyrészt az individuumra, másrészt pedig az állampolgárra hagyja. A filozófiának nem az a feladata, hogy általános érvényű elvekből próbálja levezetni a cselekvés alapjául szolgáló döntéseket, hanem hogy a pusztá ész határain belül megkísérelje elgondolni a lét és a valóság természetét, rávilágítva egyszersmind azokra a lehetőségekre is, amelyek mellett a cselekvés alapjául szolgáló döntéseinkkel elkötelezhetjük magunkat.

## IRODALOM:

- Bernet, Rudolf 2004. *Conscience et existence. Perspectives phénoménologiques*. Paris, PUF.
- Heidegger, Martin 1979. *Prolegomena zur Geschichte des Zeitbegriffs* (GA 20.) Frankfurt a. M., V. Klostermann.
- Husserl, Edmund 1973. *Zur Phänomenologie der Intersubjektivität III*. 1929–1935. (Husserliana XV.) Szerk. I. Kern. Den Haag, M. Nijhoff.
- Husserl, Edmund 1971. *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie III*. (Husserliana V.) Szerk. Marly Biemel. Den Haag, M. Nijhoff.
- Husserl, Edmund 1976. *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie I*. (Husserliana III/1.) Szerk. K. Schuhmann. Den Haag, M. Nijhoff.
- Husserl, Edmund 2002. *Logische Untersuchungen. Ergänzungsband I. Entwürfe zur Umarbeitung der VI. Untersuchung und zur Vorrede für die Neuauflage der Logische Untersuchungen* (Husserliana XX/1) Szerk. U. Melle. Dordrecht/Boston/London, Kluwer.
- Husserl, Edmund 2003: *Transzendentaler Idealismus. Texte aus dem Nachlass (1908–1921)* (Husserliana XXXVI.) Szerk. Robin D. Rollinger – R. Sowa. Dordrecht/Boston/London, Kluwer.



## Filozófiai naturalizmus és természeti fajták

### I. BEVEZETÉS

Természetesnek tűnhet, hogy azok a filozófusok, akik elkötelezettek a filozófiai naturalizmus valamilyen formája mellett, nem kétkelhetnek a természeti fajták létezésében sem. Ez azonban érdekes módon nincs így. De még azok sem feltétlenül értenek egyet abban, mi is egy természeti fajta, akik egyébként nem kétkelnek a létezésükben. Márpedig az a kérdés, hogy léteznek-e természeti fajták, nemcsak önmagában érdekes, hanem azért is, mert a válasznak fontos következményei lehetnek sok más a naturalistát – sőt a filozófust általában – érdeklő probléma tekintetében, mint amilyen a fizikalizmus, az okság, az indukció, vagy akár a jelentés filozófiai problémája.

Úgy tűnik számomra, hogy a kortárs naturalisták bizonytalankodása a természeti fajták létét, illetve mibenlétét illetően a modern filozófiai naturalizmus két forrásának konfliktusából ered. A naturalizmus egyik forrása az a filozófia és a tudomány viszonyára vonatkozó meggyőződés, amely szerint azt, hogy mi létezik, a tudomány hivatott eldönteni. A filozófia feladata részben a tudomány által és a mindennapi életben tételezett létezők közti viszony megértése, részben pedig az, hogy reflektáljon a tudományban sokszor reflektálatlanul alkalmazott gyakorlatra. Hiszen a tudósok sokszor beszélnek okokról, természeti törvényekről, magyarázatokról és valószínűségekről anélkül, hogy részletesebben vizsgálnák e terminusok pontos jelentését.

Kétségtelen azonban, hogy a modern tudományok szinte mindegyike, a legelvontabb fizika éppúgy, ahogyan a „speciális tudományok”, mondjuk a kémia, biológia vagy geológia, osztályozza az általa vizsgált jelenségeket. Ráadásul e tudományok feltételezik, hogy az általuk használt klasszifikáció – például a részecskék neutrínókra és elektronokra osztása, a víz és a higany közti különbség, vagy a tigrisek és a szamarak megkülönböztetése – nem szarta konvenció eredménye. Radikálisan különbözik tehát attól, ahogyan az emberi lényeket osztályozzuk állampolgárságuk, vagy ahogyan a kémia elemeket és az állatfajtákat kereskedelmi értékük szerint. A természeti fajta pedig éppen ezt jelenti: a természeti fajta az individuumok vagy (anyagtípusok) nem önkényes, természeti tényeken alapuló klasszifikációja.

A naturalizmus másik értelmezése szerint viszont a naturalizmus a radikális empirizmus egyik formája. A tudományoknak azért van kiemelkedő jelentőségük a naturalista számára, mert az empirikusan tesztelhető (és sok esetben valóban tesztelt) ismeretek mintája a természettudomány. Az empirizmus keretében azonban – mint ez a 17. századi kezdetektől fogva világossá vált – a természeti fajták fogalma nehezen értelmezhető. Ennek oka az, hogy az empirista hagyomány – itt most nem tárgyalható, de kétségkívül érdekes okoknál fogva<sup>1</sup> – szkeptikus a *természeti modalitásokkal* kapcsolatban. Márpedig a természeti fajtákba sorolás, úgy tűnik, feltételezi a modalitást. Valami ugyanis azon tulajdonságai alapján sorolható egy természeti fajtába, amelyekkel a fajtába tartozás esetén *rendelkeznie kell*. Ezeket a tulajdonságokat az arisztotelészi eredetű fogalom- és skolasztikus eredetű szóhasználatot követve *esszenciális* vagy – a nyelvújítás korára emlékeztető szóval – *lényegi* tulajdonságoknak szokás nevezni. Röviden: csak akkor lehetségesek természeti fajták, ha vannak fajta-esszenciák, de természeti esszenciákról csak akkor van értelme beszélni, ha létezik természeti modalitás is. Miután azonban utóbbi léte megkérdőjelezhető, az sem igazolható, hogy létezik nem önkényes, nem pusztán konvención alapuló klasszifikáció, tehát léteznek természeti fajták.

A naturalista zavarát a természeti fajtákkal kapcsolatban kiválóan példázza Quine (a filozófiai naturalizmus vitathatatlan korifeusa) jól érzékelhető bizonytalankodása a témában. Quine kíméletlen (az empirista hagyományon belül talán a legkíméletlenebb) kritikusa a *de re*, tehát nem pusztán a jelentésen alapuló „metafizikai” vagy természeti modalitás eszméjének. (Vö. Quine 1953/2002.) Kis túlzással véleménye úgy foglalható össze, hogy komolyan gondolkodó filozófus nem hihet az efféle metafizikai képtelenségekben, amelyek elkerülhetetlenül logikai ellentmondásokhoz vezetnek. Amikor azonban a természeti fajták kérdésére kerül a sor, érzékelhetően elbizonytalanodik.

Nemcsak arról van szó, hogy minden tudomány klasszifikál. Az osztályozás önmagában ugyanis még tekinthető akár konvencionálisnak is. A probléma gyökere abban áll – ahogyan azt egyébként már Locke is felismerte<sup>2</sup> –, hogy a tudomány nem csak közvetlenül észlelhető, ún. fenomenális tulajdonságok alapján klasszifikál. Egy empirista számára azonban magyarázatra szorul, hogy logikailag miként függnek össze a tapasztalatban közvetlenül megnyilvánuló fenomenális tulajdonságok a tudományok számára alapvetően fontos, de nem fenomenális „teoretikus”, vagy általában diszpozicionális fogalmakkal. A természeti fajtába

<sup>1</sup> Ezek egy részét részletesen taglalom *Modern metafizika* című könyvemben. Vö. Huoranszki 2001. 124–130.

<sup>2</sup> „Elvégre az rendelkezik a legtökéletesebb ideával a szubsztanciák sajátos fajtáinak bármelyikéről, aki a legtöbb egyszerű ideát gyűjtötte össze és szerkesztette egybe azok közül, melyek az illető szubsztanciában léteznek, amelyek közé be kell sorolnunk a szubsztancia aktív és passzív képességeit is; mert bár ezek nem egyszerű ideák, de e tekintetben, az egyszerűség kedvéért mégis alkalmas lesz őket azok közé számítani” (Locke 1690/2003. 326).

tartozás természetes magyarázatnak tűnik: a víz fenomenális tulajdonságai és az, hogy oldja a cukrot, azért tartoznak össze, mert mindkettő az azonos természeti fajtába tartozás tényéből következik.

A természeti fajtákra történő hivatkozás így például választ adhat az empirista naturalizmussal szembeni legnagyobb kihívására: az indukció (legalábbis egyik) problémájára. Arra a hume-i problémára tehát, hogy miként következtethetünk a fenomenális tulajdonságokból oksági vagy egyéb diszpozíciókra. Hume válasza, mint jól ismert, szkeptikus válasz: sehogyan sem. De Hume érvei épp azért tűnnek olyan meggyőzőnek, mert a természeti fajták tagadásán alapulnak. Másfelől viszont Hume érveiből a tudománnyal szembeni szkepticizmus is következik, ami pedig éppen nem a modern naturalista jellegzetes attitűdje. Ha tehát sikerül egy empirista számára elfogadható értelmezést adni a természeti fajtákról, talán lehetséges az indukció egyik problémájára is megfelelő naturalista választ találni.

## II. TERMÉSZETI FAJTÁK ÉS ESSZENCIÁK

De mik is a természeti fajták? Úgy tűnik számomra, hogy ez a naturalista metafizika egyik alapkérdése. Az első probléma, amit tisztáznunk kell, hogy a természeti fajták tulajdonságok-e. Első látásra úgy tűnhet, hogy nyilvánvalóan azok: akárcsak más tulajdonságoknak, a természeti fajtáknak is lehetnek instanciái vagy példányai, tehát nem lehetnek partikulárék. Bizonyos tekintetben azonban a természeti fajták hasonlóan viselkednek a partikulárékhoz: például intrinzikus tulajdonságokat állíthatunk róluk. Azt mondhatná erre valaki, hogy más tulajdonságokról is állíthatunk ilyesmit. Például mondhatjuk, hogy a piros az egy szín; vagy hogy a bátorság erény. Csakhogy amikor ilyen módon állítunk valamit egy tulajdonságról, akkor nem teszünk mást, mint hogy egy logikailag magasabb rendű, determináló tulajdonság alá soroljuk. Amikor azonban azt állítjuk, hogy a jegesmedve fehér vagy az arany sárga akkor nem ezt tesszük: a természeti fajtát kifejező terminusról pont abban az értelemben állítunk valamit, ahogyan azt egy individuumból tesszük. Következésképp csábító az a gondolat, hogy a természeti fajták vagy hibrid metafizikai entitások – kicsit partikulárék, kicsit tulajdonságok –, vagy legalábbis olyan entitások, amelyekre individuumokhoz hasonló módon utalunk. Ez utóbbi ötletet némely nagy hatású szemantikai elmélet is alátámasztani látszik. Kripke például, mint jól ismert, a természeti fajták szemantikáját az individuumbólhoz hasonlítja. (Vö. Kripke 1972/2007.)

Kripke metafizikai érveivel a következőkben lesz még dolgunk, most csak annyit szeretnék tisztázni, hogy bármilyen csábosnak tűnjék is a természeti fajtákat önálló kategóriának tekinteni, szerintem azok is csak olyan tulajdonságok, amelyeket individuumok instanciálnak. Hogy miként lehetséges mégis más, logikailag nem magasabb rendű és nem determináló jellegű tulajdonságokat ál-

lítani róluk, az talán az általam a természeti fajtákról javasolt elmélet alapján érthetővé válik.

Egy másik olyan kérdés, amelyet most méltánytalanul kurtán kell kezelnem – és amely egyébként nem független az előzőkben említett kérdéstől –, az esszenciális tulajdonságokra vonatkozik. Ha a természeti fajtákról azt kérdeztük, *mit* állíthatunk róluk, úgy az esszenciális tulajdonságokkal kapcsolatban az a kérdés merül fel, hogy *miről* állíthatjuk őket. Tulajdonságok révén nyilván állíthatjuk őket partikulárékról. Ám a fentiek alapján világos, hogy állíthatjuk őket természeti fajtákról is. Mi a viszony az esszenciális tulajdonságok e két alkalmazása között?

E kérdést – mint a legtöbb hasonló jellegű kérdést – lehet szemantikai és metafizikai szempontból is vizsgálni. Úgy gondolom azonban, a szemantikai választól önmagában nem adódik metafizikai következmény. A metafizikai kérdés ugyanis az, hogy minek vannak esszenciális tulajdonságai: individuumoknak, fajtáknak vagy mindkettőnek. Mint említettem, e helyen válaszom csak kurta és érveket nélkülöző lehet. A következőkben feltételezem, hogy nincsenek individuális esszenciák, következésképp egy individuum azonosságának se nem szükséges, se nem elégséges feltétele, hogy egy meghatározott fajba tartozzék.<sup>3</sup> Következésképp, ha vannak esszenciális tulajdonságok, azok csak annyiban esszenciálisak, amennyiben valamely individuum az adott fajba tartozik. Ennek fontos következményei lesznek az általam javasolt elméletre nézve. Mint látni fogjuk, amikor egy tulajdonságot esszenciálisnak mondunk, az mindig egy fajta-specifikus jellegzetességet fejez ki. A macskáknak nem esszenciális tulajdonságuk, hogy fehérek, a jegesmedvéknek viszont az. A tulajdonságok tehát nem önmagukban, hanem egy természeti fajta viszonylatában lényegiek. Sokak szerint ez elfogadhatatlan, én azonban amellet szeretnék érvelni, hogy nincs ebben semmi kivetnivaló.

### III. TERMÉSZETI FAJTÁK ÉS DISZPOZÍCIÓK

Térjünk vissza egy pillanatra Quine elképzeléseihez. Mint láttuk, Quine és sokan őt követve a természeti fajtákban látták az indukció egyik – az oksági diszpozíciókhoz kapcsolódó – problémájának megoldását.<sup>4</sup> Eszerint a diszpozíciók és a megfigyelhető tulajdonságok nem esetleges módon járnak együtt, amennyiben a tulajdonságok mindkét csoportja azonos természeti fajtához tartozik.

<sup>3</sup> Ennek fő oka, hogy nem tartom logikailag lehetetlennek, hogy egy partikularé azonossága akkor is megmaradjon, ha megváltozik az a természeti fajta, amelybe tartozik. Erről részletesebben lásd Mellor 1977.

<sup>4</sup> Az indukció problémájának modern kifejtése, amelyre a legtöbb kortárs szerző közvetlenül hivatkozik, Goodman 1995-ben található.

De mit értünk itt természeti fajtán? Mint láttuk, Quine ezt annak a kérdésnek a keretében vizsgálja, hogy miként vagy mi alapján tulajdoníthatunk egy „teoretikus tulajdonságot” vagy diszpozíciót egy tárgynak. A quine-i – és azóta is igen népszerű – javaslat lényege, hogy a diszpozíciókat megfigyelhetetlen, ámde nem diszpozicionális, ún. strukturális tulajdonságokkal kapcsoljuk össze. Pl. amikor azt mondjuk a vékony jégről, hogy törékeny, akkor – anélkül persze, hogy tudnánk – a jég a kristályszerkezetére utalunk. Ha pedig azt, hogy a cukor oldódik a vízben, akkor a cukor molekuláris szerkezetére.<sup>5</sup>

E kapcsolat pontosabb metafizikai természete nagy és nehezen megválaszolható probléma. Quine eliminativista magukkal a diszpozíciókkal kapcsolatban: a diszpozíciók voltaképp nem tulajdonságok, csak olyan terminusok, amelyek többé-kevésbé következetes használatát az teszi lehetővé, hogy feltesszük, a fenomenálisan hasonlatos tárgyak mikrostruktúrája azonos típusba tartozik. (Quine 1969/1999. 393–395.) Mások – így például Armstrong – diszpozicionális realisták: a diszpozíciókat tulajdonságoknak tekintik, amelyek azonosak strukturális vagy „kategoriális alapjukkal”. (Armstrong 1968. 85–88.) Megint mások szerint a diszpozíciók ugyan tulajdonságok, ám a strukturális és diszpozicionális tulajdonságok közti viszony nem lehet azonosság. Például törékeny lehet a vékony jég és a kínai váza is annak ellenére, hogy más típusú molekulák alkotják őket. De ha a viszony nem azonosság, akkor mi? Egyes filozófusok egyetértenek Armstronggal abban, hogy akármi is, nem lehet metafizikailag kontingens, és ezért a ráépülés vagy szupervenienencia fogalma segítségével igyekeznek megérteni. (Vö. Prior–Pargetter–Jackson 1982.) Megint mások Locke-ot követik, és elfogadják, hogy a viszony kontingens. (Vö. Copi 1954.)

E kérdés megoldása elengedhetetlen a kortárs naturalizmus egyik leggyakrabban vitatott témájának, a fizikalizmus problémájának megértéséhez. A fizikalizmus kérdésével itt nem lesz módom foglalkozni, de a mikrofizikai és diszpozicionális tulajdonságok viszonyára a későbbiekben még röviden visszatérek. Először azonban azt szeretném megvizsgálni – és ez jelen írás fő kérdése –, hogy mit tekinthetünk és milyen alapon egy természeti fajta esszenciális tulajdonságának. Quine, amennyire tudom, erről nem nyilatkozik egyértelműen, nyilvánvalóan azért, mert bár elkerülhetetlennek tartja a naturalista számára a természeti fajták alkalmazását, a *de re* modalításokat involváló esszenciális tulajdonságokat igyekszik elkerülni. Természetesnek tűnik azonban, hogy Quine javaslatát úgy értsük, mint a locke-i értelmében vett „reális esszenciák” rehabilitációját: a tárgyak vagy anyagfajták osztályozásának metafizikai alapját nem „manifeszt”, fenomenális tulajdonságaiknak és/vagy diszpozicionális tulajdon-

<sup>5</sup> Quine szerint tehát az „*x* tárgy törékeny” állítás logikai szerkezete a következő:  $\exists y (Mxy \ \& \ y \text{ összetörök})$ , ahol „*M*” valami olyasmit fejez ki, hogy „hasonló molekuláris szerkezetű”. (Quine 1960. 224.)

ságainak valamely együttese alkotja (ezek Locke „nominális lényegei”), hanem mikrofizikai szerkezetük.

Fontos azonban látni, hogy a mikrofizikai vagy molekuláris szintű tulajdonságok csak akkor tölthetik be azt a szerepet, amelyet Quine eredetileg szánt nekik, ha ezek egyben *magyarázzák* is a megfigyelhető és diszpozicionális tulajdonságokat. Hiszen azért vezettük be a mikrofizikai tulajdonságok segítségével azonosított „reális esszenciákat”, hogy megértsük, miként lehetséges megfigyelhető tulajdonságok alapján „teoretikus tulajdonságokat” vagy diszpozíciókat tulajdonítani. Mint látni fogjuk azonban, a természeti fajták ezen értelmezésének egy súlyos nehézséggel kell szembenéznie. Amit Quine igyekezett figyelmen kívül hagyni, az az, hogy a természeti fajták azonosításának modális következményei vannak.

Nem arról van szó, hogy a mikrostruktúra valamilyen értelemben ne magyarázhatná a fenomenális tulajdonságokat még akkor is, ha tagadjuk a *de re* modalitásokat, bár természetesen e magyarázat lehetősége sem problémamentes. Az igazi probléma véleményem szerint abból fakad, hogy csak akkor van értelme természeti fajtákról beszélni, ha azt is feltesszük, hogy valami akkor tartozik egy adott fajtába, ha rendelkezik bizonyos tulajdonságokkal. Úgy tűnik tehát, hogy az egyedek, illetve az anyagok természeti fajtákba sorolásának vannak *szükséges* feltételei. Ha ezek a szükséges feltételek „természetiek”, tehát nem pusztán konvenció alapján illetik meg a fajtába tartozó egyedeket, akkor a fajtába sorolás *de re* modalitást feltételez. És éppen ez az, amit hagyományosan úgy szokás kifejezni, hogy bizonyos tulajdonságok esszenciálisak vagy lényegiek.

#### IV. ESSZENCIÁLIS TULAJDONSÁGOK ÉS DEFINÍCIÓK

De melyek ezek az esszenciális tulajdonságok? Természetesen a filozófustól nem várhatjuk el, hogy felsorolja ezeket, már csak azért sem, mert mint korábban említettem, az esszenciális tulajdonságok fajta-specifikusak, tehát rendkívül sokfélék lehetnek. Vannak azonban olyan e tulajdonságokra vonatkozó ismertelméleti és metafizikai jellegű kérdések is, amelyek filozófiai vizsgálódás tárgyát is képezhetik.

Tegyük föl, hogy valamely természetes fajta esszenciális tulajdonságai azok, amelyeket egy szótári definícióban találhatunk. Egy gyakran használt példát alapul véve, a tigris nagy testű ragadozó, négy lábú macskaféle, bundája sárgásbarna, stb. Vajon helyes-e úgy gondolni ezekre a tulajdonságokra, mint amelyek szükségesek ahhoz, hogy valami tigris legyen, méghozzá abban az értelemben, hogy a fentiek a tigrislét esszenciális jellemzői? Amellett szeretnék érvelni, hogy a helyes válasz: igen. Méghozzá – és mint látni fogjuk, ez alapvető metafizikai jelentőséggel bíró állítás – egy kvalifikáció nélküli igen: ezek és csakis ezek a tu-

lajdonságok a lényegiek.<sup>6</sup> Azok tehát, amelyeket a fajtába tartozás *analitikus módon implikál*. Kevés filozófus fogadná ezt el manapság. De mint Kanttól tudjuk, a filozófiai igazság ügyében nem mindig tanácsos az általános véleményt követni.

A legnyilvánvalóbb ellenvetés azzal szemben, hogy a szótári definíció jellemző tulajdonságokat esszenciálisnak tekintsük, az empirizmus egyik dogmájából származik. Eszerint egy szótári definíció csak azt határozza meg, miként használunk egy kifejezést, mondjuk a „tigris” szót. Semmit sem mond azonban arról, valójában mi is egy tigris, mert nem a tigrisről, hanem a szavak jelentéséről szól. Ennek alapján talán mondhatjuk, hogy a szótári definíciók a Locke által „nominális lényegeknél” nevezett tulajdonságokat azonosítják. De ha egy esszencia „csak nominális”, akkor nem lehet *de re* szükségszerű. Következésképp nem lehet alapja egy természeti – tehát például nyelvi konvenciótól független – kapcsolatoknak a fajta és bizonyos tulajdonságok között.

Véleményem szerint ez az érv egy súlyos félreértésen alapszik: nevezetesen azon, hogy egy szótári definíció helyes vagy helytelen voltának nincsenek *a posteriori* feltételei. Ezt azonban semmi sem támasztja alá. Ki használna ma egy olyan szótárt, amelyben az szerepel, hogy a víz nem vegyület, hanem a négy elem egyike, vagy amelyik úgy jellemezné a delfint, hogy az egy hal? A szótári definíciók tudásunk gyarapodása fényében változnak. Ráadásul egy fajmeghatározásra vagy a kémiai elemek és vegyületek meghatározására szolgáló szótár sokkal specifikusabb leírást fog adni egy természeti fajtáról, mint mondjuk a Magyar Értelmező Szótár. És nyilvánvaló, hogy az e meghatározásokban szereplő tulajdonságok jó része tudományos felfedezés eredményeként került a szótárba.

Röviden: a természeti fajták szótári definíciói, meghatározásai (mint az életben annyi más) lehetnek jók és rosszak. Hogy jók-e, az bonyolult kérdés, amelynek számos a posteriori, empirikusan felfedezhető aspektusa van. Félreértés azt gondolni, hogy azért, mert egy szótár alapvető funkciója a szavak jelentésének azonosítása, a szótári definíció *csak* a jelentésekről szól. Az értelmező szótárból, ha jó, pont úgy tanulhatunk a természeti fajtákról, ahogyan egy biológia fajmeghatározóból, ami végső soron csak egy sokkal jobb (mert sokkal több információt tartalmazó) szótár. Következésképp semmi okunk arra, hogy a szótári definíciót pusztán „nominálisnak” tartsuk.

De miként lehetséges azt állítani, hogy a fajta-esszenciák olyan tulajdonságok, amelyeket a fajtába tartozás analitikus módon implikál, és amelyeket egy történetileg változó szótári jelentés ad vagy adhat meg? Talán meglepően hangzik, de én ebben nem látok semmi kivetnivalót. Analitikus következmény az, ami egy fogalom jelentéséből adódik. Egy fogalom jelentése azonban megváltozhat

<sup>6</sup> Kvalifikáció nélkül, amennyiben ezek a tulajdonságok nem pusztán „nominális” esszenciák. Mint látni fogjuk, az általam javasolt elméletből az következik, hogy nem tehető értelmes különbség „nominális” és „reális” esszenciák közt.

anélkül is, hogy ez szükségképp a referencia változását vonná maga után. Állításom az, hogy amikor felfedeztük, hogy a víz nem elem, akkor ezzel a víz jelentése is módosult. Ebből az állításból két talán ijesztőnek hangzó következmény azonban *nem* adódik. Mindenekelőtt, nem adódik, hogy *bármi*, amit felfedezünk a vízről, módosítaná a jelentését. Ha felfedezzük, hogy a Marson is van víz, abból semmi sem következik a víz jelentésére nézve. Másodszor nem adódik, hogy a jelentés változásával a terminus referenciája is megváltozna. Abból, hogy két kifejezés jelentése különbözik, nem következik, hogy referenciájuk is eltér. Aki a vizet a négy elem egyikének tartja, és úgy véli, hogy nem részecskékből áll, az nem pontosan ugyanazt érti a „víz” kifejezésen, mint aki szerint a vizet H<sub>2</sub>O molekulák alkotják. De ebből még nem következik, hogy ne referálhatna ugyanarra a természeti fajtára, mint mi.

Állításom valódi következménye inkább az, hogy helytelen az analitikus implikáció fogalmát kizárólag a „csak akkor aggregény, ha nőtlen” típusú kijelentések mintájára értelmezni. Nyilvánvaló, hogy a pusztán társadalmi konvención alapuló fogalmak jelentését nem befolyásolja semmilyen empirikus felfedezés. Még ha minden aggregény kopasz lenne is, abból sem következne, hogy a „csak akkor aggregény, ha kopasz” kijelentés analitikus igazság. De semmi sem zárja ki, hogy egy természeti fajta jelentését befolyásolják bizonyos, a fajtára vonatkozó felfedezések.

## V. ESSZENCIÁLIS TULAJDONSÁGOK ÉS DISZPOZÍCIÓK

Úgy tűnik azonban, van egy sokkal fogósabb, logikai természetű érv is az ellen, hogy a szótári definíciókban szereplő (vagy ilyen jellegű) tulajdonságokat fajta-specifikus esszenciáknak tekintsük. Ez pedig a következő. A lényegi tulajdonságok szükségszerű módon illetik meg az azonos fajtába tartozó egyedeket. Ebből pedig az látszik következni, hogy ha valami híján van egy fajta-specifikus esszenciális tulajdonságnak, az nem is tartozhat az adott fajtába. Például, ha a négylábúság valóban esszenciális tulajdonsága a tigrisnek, akkor a „háromlábú tigris” *contradictio in adjecto* kellene, hogy legyen. Ezért ha egy tigris elveszítené az egyik lábát, rögvest egy másik fajhoz kellene, hogy tartozzék. Ezt azonban senki nem gondolja így. Következésképp a négylábúság nem lényegi tulajdonság. (Kripke 1972/2007. 93.) A szótári definíció *sztereotípiákat*, nem pedig lényegi tulajdonságokat ad meg. Ennek pedig legföljebb szemantikai vagy episztemológiai jelentősége van, metafizikai aligha.

Őszintén meg kell vallanom, meglepő számomra, hogy ezt az érvet milyen sokan – és milyen jelentős filozófusok – tartják fontosnak, sőt akár döntőnek is az esszenciális tulajdonságok kapcsán. Pedig ha belegondolunk, mi következik belőle, legalábbis súlyos kétségek merülnek fel elfogadhatóságával kapcsolatban. Valóban úgy gondoljuk: az, hogy a tigrisnek hány lába van, éppoly esetleges,



mint az, hogy pontosan milyen árnyalatú a csík a háta közepén? Általában: egy biológiai faj végtagjainak száma csak a véletlen műve? Úgy esett, hogy a legtöbb tigris lábának száma négy, de lehetne persze három vagy akár hét is? Ez teljesen abszurd elképzelésnek tűnik számomra. Következésképp a példa, ha bizonyít is valamit, az nem az, hogy a szótári definíciókban szereplő tulajdonságok nem lehetnek esszenciálisak, hanem valami egészen más. Először azt próbálom tisztázni, miért *nem* bizonyítja a példa, hogy a szótári definíció jellegű tulajdonságok nem lehetnek esszenciálisak. Azután igyekszem megmutatni, valójában mit bizonyít.

Gondoljunk a tigris egy másik szótári definíció jellegű tulajdonságra, nevezetesen arra, hogy ragadozó. Ez a tulajdonság egy viselkedési diszpozíció: arra utal, hogyan táplálkozik a tigris, illetve hogy milyen módon szerzi meg a táplálékát. Abból, hogy a tigris ragadozó, nem következik, hogy folyamatosan ragadozik. Valójában a tigrisek többsége nagyon kevés időt tölt vadászattal és táplálkozással. És az is lehetséges, hogy van olyan egyed, amelyik sohasem vadászik, például mert fiatalon kimúlik. Ettől azonban még ragadozó: azokban a pillanatokban is, amikor nem vadászik, és még akkor is, ha sohasem vadászott. Ragadozónak lenni diszpozíciós tulajdonság, amely csak bizonyos körülmények között manifesztálódik (éhség, áldozat jelenléte stb.). Egy ilyen tulajdonság azonban nem szűnik meg létezni pusztán azért, mert a körülmények esetleges módon nem teszik lehetővé azt, hogy manifesztálódjék.

Nem minden diszpozíció viselkedési diszpozíció azonban. Számos olyan diszpozíció létezik, amelyek manifesztációi *állapotok*, nem pedig események. Az emberi test hőmérséklete például konstans, szemben a hullókével. Ez azt jelenti, hogy mindnyájan lényegileg diszponálva vagyunk arra, hogy egy bizonyos testhőmérsékleten éljük le az életünket. Mármost, amikor azt állítjuk, hogy a tigris lényegi tulajdonsága az, hogy négylábú, akkor éppen úgy nem azt állítjuk, hogy minden tigrisnek négy lába van, mint ahogyan akkor, amikor azt állítjuk, hogy ragadozó, nem azt akarjuk mondani, hogy állandóan vadászik; vagy amikor azt állítjuk, hogy testhőmérsékletünk konstans, nem azt akarjuk mondani, hogy senki se lehet lázas.

A négylábúság azért esszenciális tulajdonsága a tigrisnek, mert minden tigris *születésénél fogva diszponált* a négylábúságra. És ez természetesen igaz lehet akkor is, ha netán elveszíti az egyik lábát. A diszpozíciókat manifesztációik segítségével azonosítjuk: a ragadozó jelleget a vadászattal, a négylábúságot a négy láb birtoklásával.<sup>7</sup> Metafizikai értelemben esszenciális tulajdonságoknak pedig vagy e diszpozíciókat, vagy manifesztációikat szokás tekinteni: ez az, amire a szótári definíció utal. Ha például egy hulló egész életén át egy olyan helyiségben tartózkodna, amelynek hőmérséklete konstans, az aligha bizonyítaná, hogy a hulló lényegi tulajdonsága a konstans testhőmérséklet; és pontosan azért nem bizo-

<sup>7</sup> Erről bővebben lásd Huoranszki 2008.

nyítaná, mert az állandó testhőmérséklet ebben az esetben nem a hüllő diszpozíciójának megnyilvánulása. A hüllő diszpozíciójának megnyilvánulása az, hogy testhőmérséklete alkalmazkodik a környezet hőmérsékletéhez.

Ebből természetesen nem következik, hogy minden tulajdonság, amit egy szótárban találunk, valóban esszenciális. Hiszen kiderülhet, hogy tévesen azonosítottunk egy esszenciálisnak vélt tulajdonságot. Más kérdés, hogy sok *tényleges* példát találhatunk-e egy ilyen tévedésre, ha az esszenciális tulajdonságokat viselkedési diszpozícióknak vagy a diszpozíciók megfigyelhető megnyilvánulásainak tekintjük. Talán a fekete hattyúk esete lehetett ilyen: kiderült, hogy a korábban lényeginek tekintett tulajdonság nem szükségképpen illeti meg a hattyúkat. Azon tulajdonságok túlnyomó többsége azonban, amelyek alapján a vizet víznek, a tigriseket tigrisnek stb. tekintjük, nem sokat változtak az idők során. Éppen ez teszi lehetővé, hogy a jelentés bizonyos módosulása ne vonja maga után a referencia megváltozását. Ami sokat változott és változik, az a felosztás finomsága: újabb diszpozíciók azonosításával újabb természeti fajtákat azonosíthatunk és a már azonosított fajtáknak újabb lényegi tulajdonságait fedezhetjük fel.

Az a tény, hogy az esszenciális tulajdonságok vagy diszpozíciók, vagy azok megnyilvánulásai, magyarázatot adhat a szótári definíciókkal kapcsolatos egy másik gyakran hangoztatott ellenvetésre is. Kripke szerint lehetséges, hogy valami rendelkezze „az arany tulajdonított összes azonosító jellel (minden tulajdonsággal, amely alapján az aranyat eredetileg azonosítottuk), ám ez a másik anyag nem ugyanahhoz a fajtához tartozik, mint az arany, nem ugyanaz az anyag”. (Kripke 1972/2007. 92.) Véleményem szerint ez tökéletesen lehetetlen. Igaz, a piritet, a „bolondok aranyát” sokan aranynak *nézték*. Miért? Mert figyelmen kívül hagyták azt aényt, hogy egy természeti fajta azonosításához nem elégséges a fenomenális tulajdonságok azonosítása. Az anyagok esetében – éppúgy, ahogyan a fajok esetében, ahol talán ez sokkal nyilvánvalóbb – a viselkedési diszpozíciók is lényegi tulajdonságok. Ha a piritet sósavba tesszük, és oldódni kezd, rögtön felismerjük, hogy nem arannyal van dolgunk, mert az aranynak lényegi tulajdonsága, hogy csak a királyvízben oldódik.

Végezetül két megjegyzés. Először is Kripke szerint az is elképzelhető, hogy legyen valami, ami az arany vagy a tigris minden szótári definíció jellegű tulajdonságával (legyen az fenomenális tulajdonság vagy viselkedési diszpozíció) rendelkezik, és mégsem arany vagy tigris. Úgy gondolom, ha a fenti érvek a szótári definícióban említett tulajdonságok esszenciális jellege ellen esnek, sokkal kevésbé hajlunk arra, hogy ezt a metafizikai lehetőséget elfogadjuk. Később azonban, miután tisztáztam, mit gondolok a fenti példák valódi metafizikai jelentőségéről, még részletesebben is visszatérek erre a kérdésre.

Másodszor szeretnék egy rövid diagnózist megkockáztatni azzal kapcsolatban, miért tűntek ezek az érvek meggyőzőnek oly sok filozófus számára. A probléma gyökere, mint sok más esetben is, a hume-i filozófiában keresendő. Hume egyik

legfontosabb érve az oksági szükségszerűség ellen – ami számára a természeti szükségszerűség egyetlen lehetséges formája – ugyanis éppen az, hogy a természeti fajták csupán kontingens módon kapcsolódnak az oksági-viselkedési diszpozíciókhoz. A kenyér ma táplál, holnap talán nem. (Hume 1748/1973. 51.) Lehetséges-ez? Igen, ha a kenyérnek (pontosabban alkotóelemének, a gabonának, mert ez a természeti fajta) nem lényegi tulajdonsága az, hogy minket táplál. Márpedig Hume szerint nem lehet az, mert a tápértékről nincs „benyomásunk”, természeti fajtákat pedig csak benyomások alapján azonosíthatunk. Tehát: a bolondok aranya talán tényleg arany – csak furcsa diszpozíciókkal rendelkezik –, az oksági diszpozíciókra történő következtetés pedig lehetetlen. Ha azonban – Locke-ot követve – nem látjuk indokoltnak az esszenciális tulajdonságok efféle korlátozását, nyilvánvaló, hogy a mérgező búza nem búza, a bolondok aranya nem arany. Végül is pontosan tudjuk, hogy csak a bolond tartja aranyának, mert csak egy bolond lehet meggyőződve arról, hogy a dolgok szükségképpen azok, amiknek felszíni, fenomenális tulajdonságaik alapján tűnnek.

## VI. ESSZENCIÁLIS TULAJDONSÁGOK ÉS *DE RE* SZÜKSÉGSZERŰSÉG

Mindeddig keveset szóltam arról, mit értünk *de re* modalitásokon, márpedig ha a fajták valóban természetiek, akkor a fajta-esszenciák *de re* értelemben szükséges feltételei a valamely fajtra tartozásnak. A Kripke által korábban tárgyalt ellenvetés megfogalmazható más formában, a modalitások sztenderd nyelvén is. Eszerint egy tulajdonság csak akkor lehet esszenciális, ha minden lehetséges világban, amelyben az adott fajta létezik, valami csak akkor tartozhat az adott fajtához, ha a megfelelő tulajdonsággal is rendelkezik. Ha tehát a négylábúság a tigrisek esszenciális tulajdonsága, akkor valami csak akkor lehet tigris, ha minden lehetséges világban négylábú. Másképp fogalmazva, ha lehetséges, hogy valami, ami négylábú, háromlábú legyen, és mégse veszítse el tigris voltát, akkor a négylábúság nem lehet az adott fajta esszenciális tulajdonsága. A példa természetesen bármely más eddig említett tulajdonságra is kiterjeszthető. Következésképp a szótári definíciókban említett tulajdonságok nem lehetnek lényegiek.

Az egyik lehetséges válasz erre az ellenvetésre ismét csak az, hogy a lényegi tulajdonságok diszpozíciók, és a diszpozíció nem vesz el még akkor sem, ha kontingens okoknál fog nem manifesztálódik. Sajnos azonban, ez önmagában nem elegendő a modális probléma megoldásához. Továbbra is válaszra vár, miként lehetséges, hogy valami szükségképpen rendelkezék a négylábúság diszpozíciójával, ámde *lehetséges*, hogy csak három lába legyen. A diszpozíciókat manifesztációik segítségével azonosítjuk: akkor miként rendelkezhet valami egy diszpozícióval anélkül, hogy – a megfelelő körülmények között – a diszpozíció mibenlétét azonosító esemény vagy állapot megvalósulna? Ezzel a kérdéssel ak-

kor is szembe kell néznünk, ha – helyesen – a lényegi tulajdonságoknak a diszpozíciókat tekintjük.

Ámde éppen ez az a kérdés, amely véleményem szerint a természeti fajtákat érintő metafizikai probléma gyökeréhez vezet. Ha a természeti fajtákba sorolás feltételezi az esszenciális tulajdonságokat, az esszenciális tulajdonságok pedig modalitásokat implikálnak, ezek a modalitások *természeti* modalitások kell, hogy legyenek. Hogyan értsük ezeket? Kripke és a kortárs filozófusok közül szinte mindenki a természeti modalitásokat (is) *a lehetőségekre kiterjesztett kvantifikáció* segítségével igyekszik megérteni.<sup>8</sup> Ha szükségszerű, hogy a tigrisek négylábúak, akkor minden lehetséges világban minden, ami tigris, az négylábú. Ha tehát lehetséges, hogy valami tigris is legyen, meg háromlábú is – azaz van olyan lehetséges világ, amelyben egy tigris háromlábú –, akkor a négylábúság nem lehet a tigrisek lényegi tulajdonsága. Dióhéjban ez Kripke és az őt követő filozófusok érve mellett, hogy a szótári definícióban említett tulajdonságok nem lehetnek lényegiek.

Természetesen nem csak ezek az alternatívák adottak számunkra. Először is választhatjuk a quine-i stratégiát, és kísérletet tehetünk arra, hogy a természeti fajtákat esszenciális tulajdonságokra történő hivatkozás nélkül értelmezzük. Paradox módon Kripke pontosan ezt teszi. Hagyományosan, mint említettem, az esszenciális tulajdonságok külön-külön szükséges, de nem elégséges feltételei annak, hogy valamely egyed vagy anyag egy fajtához tartozzék. Mint látni fogjuk, Kripke elméletéből az következik, hogy az esszenciális tulajdonság nemcsak szükséges, de *elégséges* feltétele is a valamely fajtába tartozásnak. Mégpedig azért, mert a fajtába tartozás voltaképp *ugyanaz a tulajdonság*, mint a fajta lényege. Amikor fajtába sorolunk valamit, és amikor azt állítjuk, hogy van egy esszenciális tulajdonsága, végső soron ugyanarra a tulajdonságra utalunk, csak más-más leírás mellett. Egy tulajdonság csak akkor lényegi, ha metafizikailag szükségszerű, és csak akkor szükségszerű metafizikailag, ha *azonos* a természeti fajtával.

Mindez azt eredményezi, hogy a szó hagyományos értelmében nincsenek esszenciális tulajdonságok, olyan tulajdonságok tehát, amelyek szükségesek, de önmagukban nem elégségesek ahhoz, hogy valami egy természeti fajtába tartozzék. De vajon miért is ne tagadhatnánk, hogy léteznek esszenciális tulajdonságok ebben a hagyományos értelemben? Hiszen láttuk, már Quine sem hivatkozik rájuk. A probléma csak az, hogy Quine, kénytelen-kelletlen bár, de mégis elismerte a természeti fajták létezését, mivel ezt tartotta az egyetlen lehetséges

<sup>8</sup> A különböző modális szemantikák ezt különböző módon teszik. Lewis szerint a szükségszerűségekre és lehetőségekre vonatkozó kijelentéseinket át kell fogalmaznunk olyan kijelentésekre, amelyek explicit módon kvantifikálnak lehetséges világok felett. Kripke szemantikájában a kvantifikáció a szemantikai interpretáció szintjén jelenik meg. A kvantifikáció és a modern modális szemantika kapcsolatának metafizikai aspektusait részletesen tárgyalom a *Modern metafizika* című könyvem IV. fejezetében.

módnak arra, hogy megértsük a diszpozíciók és a fenomenális tulajdonságok kapcsolatát, amely szinte minden tudományos tevékenység és ismeret alapja. Ha viszont mindezeket a fajtába tartozás szempontjából tökéletesen esetlegesnek deklaráltuk, akkor sajnos semmit sem értettünk meg. Következésképp Kripke megoldása mintha a kád vízzel együtt a gyereket is kiöntötte volna. Van ugyan egy elméletünk a természeti fajtákról, amely kiválóan illeszkedik a modalitásokról alkotott nézeteinkhez, de eközben elvész a természeti fajtákra történő hivatkozás minden magyarázóereje.

Lehetséges azonban más megoldás is, amit a filozófusok többsége ma nem vizsgál, pedig, úgy gondolom, ebben rejlik a rejtély kulcsa. Állításom a következő: az eddig vizsgált problémák nem azt bizonyítják, hogy a szótári definíció értelmében vett tulajdonságok ne lennének (vagy legalábbis lehetnének, ha helyesen ismerjük fel őket) esszenciálisak, se nem azt, hogy a természeti fajtákat a tulajdonságazonosság segítségével kellene értelmeznünk, hanem azt, hogy *a modalitás sztenderd értelmezése alkalmatlan a természeti modalitás megértésére*. A természeti modalitás nem érthető meg a kvantifikáció segítségével, mivel nem arra vonatkozik, hogy hány – mind vagy csak néhány aktuális és lehetséges – egyed vagy partikuláré rendelkezik bizonyos tulajdonságokkal, hanem hogy *milyen módon* rendelkeznek ezekkel a tulajdonságokkal, *ha* rendelkeznek velük. Absztrakt entitások esetében a szükségszerűség azért érthető jól a kvantifikáció segítségével, mert az a kérdés, hogy *miként* instanciál egy absztrakt partikuláré (ha egyáltalán van ilyen) egy tulajdonságot, értelmetlen. Nincs értelme feltenni a kérdést, hogy a három milyen módon instanciálja a páratlanság tulajdonságát. De ez a kérdés nemcsak hogy nem értelmetlen a természeti (itt: nem absztrakt) egyedekkel és tulajdonságaikkal kapcsolatban, de elengedhetetlen részét alkotja a természeti világ megértésének.

A természeti modalitást tehát nem a kvantifikáció, hanem a határozói módosítás mintájára kell értelmeznünk. A következőkben ezt a javaslatot és ennek a természeti fajtákra és esszenciális tulajdonságokra vonatkozó következményeit fogom tárgyalni.

## VII. ESETLEGESSÉG

Mit jelent az, hogy egy tulajdonság szükségszerűen illet meg egy bizonyos fajhoz tartozó egyedet? A leglemibb értelmezés szerint a következőt: az, hogy az adott egyed a vizsgált tulajdonsággal rendelkezik, *nem pusztán kontingens* tény. A kontingencia vagy metafizikai esetlegesség fogalma azonban két dologra is utal, amelyek nem minden kontextusban fedik át egymást. Az egyik, hogy valami *nem szükségszerűen igaz*. A másik, hogy *véletlen*. Amellett szeretnék kardszkodni, hogy a természeti szükségszerűség fogalmát úgy érthetjük meg a legjobban, ha azt a *nem véletlen* szinonimájaként értjük. A természeti szükségszerűség két

alapvető formája az okság, illetve a fajta-specifikus esszenciális tulajdonságok jelenléte. Eszerint valamely okozat abban az értelemben szükségszerű, hogy bekövetkezése nem volt véletlen. Egy tulajdonság pedig abban az értelemben esszenciális, hogy nem véletlen, hogy valami rendelkezik vele.

Fontos látni, hogy az, valami véletlen-e vagy sem, lényegileg relatív fogalom. Az okozat nem véletlen esemény, feltéve az okokat. Az, hogy valami egy esszenciális tulajdonsággal rendelkezik, nem véletlenszerű, feltéve, hogy bizonyos fajta-ba tartozik. Továbbá a „véletlenszerű” legjobb értelmezése nyilvánvalóan egy határozói módosítás: valami véletlenül történt, vagy valami véletlenül valamilyen. Ezeket a kifejezéseket hajlamosak vagyunk episztémikusan értelmezni: a „véletlenül” ebben az esetben azt jelentené, hogy számunkra előre nem látható módon, tehát váratlanul.<sup>9</sup> De semmi sem kényszerít arra, hogy így értelmezzük őket.<sup>10</sup> A véletlen, akárcsak a lehetőség vagy a valószínűség, lehet metafizikai. Pontosan azt fejezi ki, mi módon történt egy esemény, illetve milyen módon instanciál valamely egyed egy tulajdonságot. És éppen ez az, amire szükségünk van a természeti modalitások megértéséhez.

Miután a „véletlen” kifejezést elsődleges értelemben eseményekre alkalmazzuk, kiindulópontom az a kérdés lesz, mit értünk azon, hogy valami véletlenül történt vagy véletlenül nem történt meg. Természetesen a véletlenszerűséget számos módon értelmezhetjük, attól függően, hogy mely kontextusban milyen típusú eseményeket vizsgálunk. Jelen kontextusban azonban viszonylag egyszerű dolgunk van: az eseményeket úgy kell vizsgálnunk, mint diszpozíciók manifesztációit. Amikor egy törekeny poharat megütök egy kemény tárggyal, az darabokra hullik: ez az esemény a törekenység manifesztációja. A törekenység tulajdonítása és az, hogy az adott körülmények közt egy törekeny tárgy összetörik, nem véletlenszerűen tartoznak össze. És éppen ebben az értelemben szükségszerű, hogy egy törekeny tárgy összetörik, ha egy kemény felülethez csapódik.

Lehetségesek-e azonban olyan körülmények, amelyek közt a törekeny pohár annak ellenére nem törik össze, hogy megütöm egy kemény tárggyal? Természetesen lehetségesek: például akkor, ha alaposan be van csomagolva, vagy úgy, hogy a filozófusok Descartes óta kedvenc karakterét hívjuk segítségül, és feltesszük, hogy egy démon varázslatos módon megóvja. Fontos látni, hogy ezekben az esetekben nem arról van szó, hogy a diszpozíció a manifesztáció körülményei között *elvész*. A törekeny pohár nem lesz kevésbé törekeny attól, hogy becsomagolják, vagy attól, hogy egy démon – csak ő tudja, hogyan – megóvja. Inkább arról van szó, hogy lehetségesek olyan körülmények, amelyek a

<sup>9</sup> „Véletlen ugyan nincs a világon, de a történések igazi okának nem ismerete mégis olyan hatással van az értelmünkre, s olyan hitet vagy érzetet ébreszt bennünk, mintha volna.” (Hume 1748/1973. 86.)

<sup>10</sup> Az alternatív értelmezésekről lásd például Owens 1992 és Elder 2004. 96–102.

diszpozíciót *álcázzák* (Johnston 1992). A törékeny pohár tehát csak *ceteris paribus* kell, hogy összetörjön, ha egy kemény tárgy megüti. Nyilvánvaló módon, az adott körülmények között *nem véletlen*, hogy összetörik, mivel ez szükségképpen következik a pohár azon tulajdonságából, hogy törékeny. Ez azonban – mint láttuk – nem zárja ki, hogy ne törjön össze, mivel lehetséges a körülmények olyan *véletlenszerű egybeesése*, amely megakadályozza, hogy a diszpozíció megnyilvánuljon. Ettől azonban maga diszpozíció nem vész el: ha egy pohár nem lenne törékeny, értelmetlen lenne csomagolással védeni; és valószínűleg egyetlen racionális démon se pazarolná az energiáit arra, hogy egy rugalmas acélrudat igyekezzen megóvni az összetöréstől. Röviden: egy diszpozíció azért kötődik szükségképpen bizonyos eseménytípusokhoz, mert ha a diszpozícióval rendelkező tárgy meghatározott körülmények között van, és *nincsenek jelen véletlenszerű álcázó tényezők*, akkor az esemény szükségképpen bekövetkezik.

Mi következik mindebből a természeti tulajdonságokra nézve? Mindenekelőtt a fajtába tartozás szempontjából meg kell különböztetnünk a tulajdonságok két csoportját. A tulajdonságok egyik csoportját a természeti, de nem esszenciális tulajdonságok alkotják. Az például, hogy a csík a hátán pontosan milyen árnyalatú, vagy az, hogy az állat pontosan milyen súlyú, bizonyos határok közt a tigris természeti és nem véletlenszerű tulajdonsága. A természeti tulajdonságok is diszpozíciók manifesztációi, de természetesen a fajtába tartozást nem befolyásolja, ha az adott egyed *nem* rendelkezik velük. Következésképp *hiányukat* a fajtába tartozás szempontjából nem tekinthetjük véletlenszerűnek. Az esszenciális tulajdonságokat mármint éppen az különbözteti meg a pusztán természeti tulajdonságoktól, hogy amennyiben egy fajta-specifikus lényegi tulajdonság *nem* manifesztálódik, az *mindig* véletlenszerű. Tehát, csak akkor lehet egy tulajdonság esszenciális, ha manifesztációjának hiányát véletlenszerű faktorok jelenléte magyarázza.

A háromlábú tigris – kicsit unalmas – példájánál maradva: a lábát elvesztő tigrisre is igaz, hogy születésénél fogva arra diszponált, hogy négy lába legyen. Véletlenszerű faktorok magyarázzák, hogy a négylábúság állapota nem áll fenn. Úgy is fogalmazhatunk: az, hogy háromlábú, a fajta-specifikus tulajdonság szempontjából pusztán véletlen. Az állatkertbe szállított tigris sem fog vadászni: nem azért, mert a vadászösztön hiányzik belőle, vagy azért, mert ne lenne éhes és mert ne lenne jelen a potenciális áldozat (például a látogatók); azért nem vadászik, mert a „csomagolás” – jelen esetben a ketrec – megakadályozza abban, hogy vadásszék. Röviden: az, hogy sohasem vadászik, a fajta-specifikus tulajdonság szempontjából pusztán véletlen.

Összefoglalva: az esszenciális tulajdonságok olyan diszpozíciók, amelyek az adott fajtába tartozó egyedek fenomenális állapotait és tipikus viselkedési megnyilvánulásait meghatározzák. Ez nem vonja maga után azt, hogy ha valamely egyed nincs az adott állapotban, vagy a releváns körülmények között nem viselkedik az elvárt módon, akkor már ne is tartozhatna az adott természeti fajtába.

Csak azt jelenti, hogy a lényegi tulajdonságok manifesztálódásának elmaradása – szemben a pusztán természeti, de nem esszenciális tulajdonságokéval – szükségképpen véletlenszerű. Éppen ebben az értelemben nem kontingens, hogy valamely egyed rendelkezik a fajta-specifikus esszenciális tulajdonságok megnyilvánulásaival.

## VIII. KONSTITÚCIÓ

A tulajdonságok eszerint metafizikai szempontból három csoportba oszthatók: vannak tulajdonságok, amelyek egy adott fajtába tartozó egyedek természeti, de nem esszenciális tulajdonságainak megnyilvánulásai; vannak, amelyek véletlenszerűek; és vannak, amelyek esszenciális módon illetik meg. Hogy a kategóriák közti különbség éles vagy elmosódott, az további kérdés. További kérdés az is, hogy az esszenciális tulajdonságok ezen értelmezése milyen értelemben modális. Ha valaki alapvetőnek tekintti a kiterjesztett kvantifikációt és a lehetséges világokra való hivatkozást a modalitás megértéséhez, akkor ez az interpretáció nem modális. Szerintem ugyan az értelmezés nyilvánvaló módon modális, csak a szokásostól eltérő beszámolót ad a természeti modalitásról, de végül is ezen semmi sem múlik. Ha valaki javaslatomat úgy értelmezi, hogy az az esszenciális tulajdonságok nem-modális értelmezése – bár ez számomra kissé dogmatikusnak tűnik –, önmagában azt sem kifogásolom. A kérdés az, sikerül-e magyarázania, mik és miért esszenciális tulajdonságok, valamint azt, hogy egy esszenciális tulajdonság miként magyaráz. A következőkben amellet érvelek, hogy az általam javasolt értelmezés jól megvilágítja azt is, miért esszenciális egy tulajdonság, és azt is, hogy az esszenciális tulajdonságok mit és miért magyaráznak.

Kiindulópontom a lényegi tulajdonságokról alkotott legdivatosabb elmélet, a kripkeánus konstitúció-esszencializmus lesz. A konstitúció-esszencializmus szerint egy természeti fajta azonosításának szükséges és elégséges feltétele a mikrostruktúra típusának azonosítása. Az anyagok esetében ez az atomszerkezetet, az élőlények esetében a génstruktúrát jelenti. Kripke érve e mellett az elképzelés mellett a merev jelöléssel kapcsolatos szemantikai érv. Ezzel az érveléssel most nem foglalkozom, csak annyit szeretnék megjegyezni, hogy ha az elmélet következménye elfogadhatatlan, nyilván a szemantika sem lehet tökéletes.

Az elméletnek van azonban két olyan metafizikai következménye, amit Kripke és az őt követők egyaránt elfogadnak, de amely számomra vitathatónak tűnik. Az első, hogy az elmélet szerint, ha egy individuum vagy anyagfajta egyetlen olyan megfigyelhető tulajdonsággal vagy viselkedési diszpozícióval sem rendelkezik, amelynek révén általában azonosítjuk, akkor sem kell tagadnunk, hogy a korábbi módon azonosított fajtába tartozik. A másik, hogy ha két egyed vagy anyagfajta minden megfigyelhető és viselkedésbeli tulajdonsága tekintetében azonos, még mindig lehetséges, hogy különböző természeti fajtába tartozzanak.



Miután ezek az elmélet logikai következményei, a konstitúció-esszencializmus csak akkor fogadható el, ha ezeket a következményeket is elfogadjuk. Véleményem szerint azonban erre semmi okunk sincs.

Az elmélet ezen konzekvenciáinak elfogadásából mindenekelőtt az következik, hogy az esszenciális tulajdonság *semmit nem magyaráz*. Ez az individuáló esszenciák (ha vannak ilyenek) esetében talán nem jelent problémát, a fajta-esszenciák tekintetében azonban szomorú következmény egy naturalista számára; hiszen, mint emlékszünk, a naturalista legfőbb indoka a természeti fajták elismerésére az volt, hogy ezek magyarázzák a fenomenális tulajdonságok és a diszpozíciók közti viszonyt. Ám ha a konstitúció és az anyagok vagy biológiai fajok egyéb tulajdonságai közti viszony teljesen esetleges, amint az elméletből következik, akkor képtelenség belátni, hogyan magyarázhatja a konstitúció az anyagok és fajok megfigyelhető tulajdonságait vagy azt, hogy ezek miként tartoznak össze. (Mint emlékszünk, ez volt Quine eredeti kérdése.) Sőt, ami ennél is szomorúbb következmény, episztémikus értelemben tökéletesen lehetetlenné teszi számunkra a hozzáférést ahhoz, mi milyen fajta tartozik. Hiszen ha a konstitúció tetszőleges észlelhető tulajdonságokkal és viselkedési diszpozíciókkal kapcsolható össze, akkor az általunk alkalmazható osztályozás teljesen alkalmatlan ennek eldöntésére.

Ez utóbbi, már-már abszurdnak tűnő következmény elkerülése érdekében az elmélet talán módosítható. E módosítás szerint – bár nem tudom, van-e olyan filozófus, aki ezt elfogadja – nem arról van szó, hogy egy természetes fajta összes megfigyelhető tulajdonsága véletlenszerűen kapcsolódik a konstitúcióhoz. Inkább arról, hogy vannak olyan tulajdonságok – elméleti, közvetlenül nem megfigyelhető diszpozíciók, mint a kémiai kötés vagy a DNS struktúrája –, amelyek alapján a konstitúció azonosítható, és csak ezekről állítjuk, hogy kontingens módon kapcsolódnak az egyéb észlelhető tulajdonságokhoz és diszpozíciókhoz.

Az elméletnek e módosítása számomra bizonyos tekintetben nagyon plauzibilisnek, másfelől viszont elfogadhatatlannak tűnik. A módosítás plauzibilis annyiban, amennyiben én is úgy gondolom – még ha igen sokan ezzel nem is értenek egyet –, hogy a konstitúciót meghatározó diszpozíciók és a megfigyelhető tulajdonságok közti viszony *logikai* értelemben kontingens.<sup>11</sup> Azonban teljesen elfogadhatatlannak tartom az elmélet azon következményét, hogy a természeti fajtákat pusztán a konstitúció, illetve az ahhoz kapcsolódó diszpozíciók alapján azonosítsuk.<sup>12</sup> Befejezésül amellet érvelek, hogy erre nincs jó okunk, és hogy az esszenciális tulajdonságokról általam javasolt elmélet jobb megoldást kínál a konstitúció és a megfigyelhető tulajdonságok viszonyának értelmezésére, mint a Kripke és követői által elfogadott elmélet.

<sup>11</sup> Mégpedig azért, mert a konstitúció típusát meghatározó diszpozíciók logikailag nem implikálják sem a fenomenális tulajdonságokat, sem a makroszintű diszpozíciókat. E kérdés részletesebb tárgyalását azonban más alkalomra kell halasztanom.

<sup>12</sup> Természetesen nem én vagyok az egyetlen, aki így gondolja. Lásd például Altrichter 1985/1993. 354.

## IX. TÖBBSZÖRÖS MEGVALÓSULÁS

Kiindulópontom egy csontig rágott – következésképp mindenki által ismert és ezért sommásan prezentálható – példa, Putnam esete a vízzel (vö. Putnam 1975). A példa felteszi, hogy valamely távoli bolygón létezik egy olyan anyagfajta, amely rendelkezik a víz összes észlelhető tulajdonságával (áttetsző, színtelen, szagtalan) és viselkedési diszpozíciójával (zéró Celsius-fokon fagy, 100 Celsius-fokon forr, oltja a szomjat stb.), ám nem H<sub>2</sub>O molekulák, hanem különös XYZ molekulák alkotják. Ahhoz persze, hogy erre valaki valaha is rájöhessen, az Ikerföldön található anyagnak legalább bizonyos körülmények közt másképp kell viselkednie, mint a földi víznek. De most tekintsünk el ettől: vagy fogadjuk el, hogy bár a különbségről ember nem szerezhethet tudomást, ettől még felfogható, tehát logikailag és ezért metafizikailag lehetséges; vagy tegyük fel, hogy vannak olyan diszpozíciók, amelyek csak különleges kísérleti körülmények között bizonyítják, hogy az anyagminták konstitúciója különbözik. Kérdésünk a következő: vajon az Ikerföldön található anyag a víz természeti fajtába tartozik-e?

A kripkeánus válasz egyértelmű: természetesen nem, hiszen a konstitúció típusa szükséges és elégséges feltétele annak, hogy valami egy bizonyos fajtába tartozzék. Következésképp az Ikerföldön található vízszerű anyag *nem* víz. De mi is kényszerítene bennünket ennek elfogadására? Mint már tisztáztuk, ahhoz, hogy a konstitúció-esszencializmus logikailag konzisztens elképzelés legyen, fel kell tételeznünk, hogy az alkotóelemek és a megfigyelhető tulajdonságok közti viszony teljességgel esetleges. Miért kellene akkor elfogadnunk, hogy a fajtába tartozás kritériuma *éppen és csakis* a konstitúció típusa? Véleményem szerint – eltekintve néhány perdöntőnek semmiképp sem nevezhető szemantikai kikötéstől azzal kapcsolatban, hogy melyik terminust tekintsük merev jelölőnek és melyiket nem – semmi.

Úgy gondolom, az általam javasolt elmélet sokkal plauzibilisebb magyarázatot nyújt a konstitúció kérdésére. Mint azt számos filozófus megjegyzi, a legtöbb, amit az Ikerföld példa bizonyít, hogy a konstitúció nem irreleváns a valamely természeti fajtába sorolás szempontjából. De a példa önmagában nem bizonyítja, hogy a természeti fajtát kizárólag a konstitúció határozná meg. Hasonlítsuk össze Putnam példáját a ráépülés vagy szupervenienencia és a többszörös megvalósulás kérdésével. Szinte mindenki elismeri, hogy bizonyos pszichológiai állapot típusok – például meggyőződés, szándékok, motivációk vagy érzetek – megjelenése számos különböző fizikai állapot típussal kapcsolható össze. Természetesen vitatható kérdés, hogy a pszichológiai állapotokat természeti fajtának kell-e tekintenünk.<sup>13</sup> Az azonban nem vitatható, hogy az Ikerföld-probléma

<sup>13</sup> Fodor nagy hatású érvéről, mely szerint a speciális tudományok, így a pszichológia által használt klasszifikáció természeti fajtákra vonatkozik, lásd Fodor 1974.

legalábbis analóg a pszichológiai állapotok kérdésével, mégpedig a következő okból.

A pszichológiai állapotokat részben az egyének viselkedési diszpozíciói, részben fenomenális jellegük segítségével azonosítjuk. Nem felfoghatatlan azonban, hogy két egymástól viselkedési diszpozíciók és fenomenális állapotok tekintetében megkülönböztethetetlen egyén különbözzék fizikai konstitúciójuk tekintetében. Amit a ráépülés vagy szupervenienencia nem enged meg, az csupán annyi, hogy két egyén pszichológiai tekintetben különbözzék egymástól, ha fizikai konstitúciójuk és tulajdonságaik tekintetében megkülönböztethetetlenek.<sup>14</sup>

Állításom az, hogy a természeti fajták és az ikerföldi példák esete ezzel analóg. Hiszen feltevésünk az ikerföldi vízszerű anyagról éppen az volt, hogy az semmilyen fenomenális és – legalábbis „hétköznapi szinten”, műszerek stb. nélkül – megfigyelhető tulajdonságában nem különbözik a mi jól ismert vizünk-től. Amiben különböznek, az a mikrostruktúra. De ha a mentális állapotok esetében a megvalósulás különbsége önmagában nem hajlamosít bennünket arra, hogy azt állítsuk, különböző típusú mentális állapotokról van szó, miért kellene azt gondolnunk, hogy csak azért, mert az ikerföldi vízszerű anyag konstitúciója különbözik a földiétől, már nem is víz? Miért ne mondjuk inkább azt, hogy felfedeztük, a vizet több típusú mikrofizikai összetevő is alkothatja?

Amennyire tudom, a legnépszerűbb megoldás manapság erre a problémára a – persze nagyon kifinomult és bonyolult logikát igénylő, de mégiscsak – kitérő válasz. Az úgynevezett „kétdimenziós szemantika” elképzelése szerint a konstitúció esszenciális is meg nem is, annak függvényében, hogy a tárgyat metafizikai vagy episztémikus szempontból vizsgálom. (Chalmers 1996. 56–65., Jackson 1998. 49–52.) Episztémikus szempontból a fenomenális és megfigyelhető diszpozíciókon alapuló tulajdonságok relevánsak, mert enélkül nem lennének képesek azonosítani, miről is beszélünk egyáltalán. Metafizikai szempontból viszont igazat kell adjunk azoknak, akik szerint a fajta azonosításának egyetlen kritériuma a konstitúció. Beszélni pedig beszélhetünk így is, meg úgy is; csak világossá kell tenni, hogy most éppen episztémikus vagy metafizikai kontextusban használjuk-e a kifejezést.<sup>15</sup>

Ezzel az elképzeléssel nekem nem az a gondom, hogy szisztematikusan megkettőzi a természeti fajták jelentését. A jelentés lehet kontextusfüggő, az pedig, hogy a kontextusfüggést mi mindenre akarjuk kiterjeszteni, olyan nyelvfilozófiai döntés kérdése, amelybe e helyütt bizonyosan nem akarok beleszólni.

<sup>14</sup> Ha a pszichológiai externalizmus igaz, akkor természetesen ez sem áll. Most azonban tegyük fel, hogy az egyén pszichológiai állapotait csak intrinzikus fizikai állapotai határozzák meg. Mint látni fogjuk, ennek nincs különösebb jelentősége az általam használt érvteltekintetben.

<sup>15</sup> A kétdimenziós szemantikáról lásd bővebben Ambrus 2007. 66–68.

Csak hogy ez az elképzelés számomra metafizikai szempontból elfogadhatatlan. Ugyan miért kellene azt gondolnunk: az a kérdés, milyen megfigyelhető (fenomenális és diszpozicionális) tulajdonságok lényegesek ahhoz, hogy valamit a víz természeti fajtába soroljunk, pusztán a szóhasználat episztémikus feltételeire vonatkozik, és nem mond semmit arról, hogy a dolgok *természetüknél fogva* bizonyos osztályokba tartoznak? Még ha el is tekintünk a viselkedési diszpozícióktól, ahogyan azt sokan – szerintem teljesen indokolatlanul – teszik, miért is ne tartozhatna valamely tárgy azon tulajdonsága, hogy milyen érzeteket kelt vagy kelthet bennünk, fajta-specifikus lényegi tulajdonságai közé?

Az elutasító válasz legfontosabb oka, úgy látom, az esszenciális tulajdonságok modalitásának helytelen értelmezése. Ha a fajta-specifikus esszencia olyan tulajdonság, amely minden egyes lehetséges világban meg kell, hogy illesse az adott fajtába tartozó egyedet, akkor metafizikai értelemben a nem konstitúció jellegű esszenciák szóba sem jöhetnek, hiszen mint láttuk, valami egy adott fajtába tartozhat anélkül, hogy rendelkezne a szóban forgó tulajdonsággal. Következésképp, ha egyáltalán vannak természeti fajták, akkor az osztályozás metafizikai alapját csak a konstitúció adhatja. Véleményem szerint azonban ez az érvelés hibás. Amit az azonos fenomenális tulajdonságokkal rendelkező fajok különböző mikrofizikai megvalósulásának logikai lehetősége bizonyít, az nem az, hogy a fajtákat a mikrostruktúrákkal kell azonosítanunk, hanem az, hogy a mikrostruktúra és a természeti fajták közti modalitást újra kell értelmeznünk.

A fentiekben a tulajdonságok három típusát különböztettem meg: a fajta-specifikus természetszerű, a fajta-specifikus véletlenszerű és a fajta-specifikus esszenciális tulajdonságokat. Úgy gondolom, jelen episztémikus helyzetünkben nincs okunk kételkedni abban, hogy a víz  $H_2O$  molekulákból áll (többnyire, a lehetséges finomításoktól – nehézvíz, izotópok stb. – itt most tekintsünk el) és hogy ez a víz esszenciális tulajdonsága, ami akár szótárainkban is megjelenhet. De tegyük föl, hogy egyszer csak egy távoli bolygón fölfedezzük az XYZ összetételű anyagot, ami pontosan úgy néz ki és úgy viselkedik, mint a víz (feltéve, de meg nem engedve, hogy ez lehetséges). Döntés elé kerülünk: víz akkor ez vagy nem? Szerintem a helyes döntés: igen, ez víz. De ha víz, akkor milyen viszonyban áll a víz mint természeti fajta és az azt alkotó mikrofizikai típus?

Az egyik lehetőség, hogy azt állítsuk, a viszony véletlenszerű, minden értelemben esetleges. Ez rendkívül kontra-intuitív, kétes értékű javaslat lenne. Ha másért nem, hát azért, mert a tudomány egyik nyilvánvaló célja a fajták konstitúciójának azonosítása, és ez feltételezni látszik, hogy a kapcsolat nem esetleges. Talán részben éppen ez magyarázza, miért tűnik olyan vonzónak az a válasz, hogy fajta-specifikus esszencia csak a mikrostruktúra típusa lehet. Az általam javasolt keretben persze van más lehetőség is. A másik lehetőség az, hogy a földi víz esszenciális tulajdonságának tekintem azt, hogy  $H_2O$  molekulák alkotják, és azt állítom, hogy pusztán véletlenszerű, hogy az Ikerföldön XYZ „molekulák”

konstituálják. Ez megengedhetetlenül sovíniszta gyakorlatnak tűnik: miért éppen a mi vízünk számít?

A számomra nyilvánvaló módon plauzibilis megoldás a harmadik lehetőség: a konstitúció a víz természeti tulajdonsága, és ennyiben természetesen *nem esetleges vagy véletlenszerű*, hogy a vizet  $H_2O$  vagy éppen XYZ molekulák alkotják. Aból azonban, hogy a víznek nem véletlenszerű, hanem nagyon is fajta-specifikus természeti tulajdonsága, hogy ez vagy az alkotja, önmagában még nem következik, hogy lényegi tulajdonsága is lenne. Semmi okunk arra, hogy ne tekintsük lényegi tulajdonságának mindaddig, amíg nem fedeztük fel az XYZ „molekulák” által konstituált vizet. Ha azonban kiderül(het)ne, hogy van olyan anyag, amely pontosan úgy néz ki és viselkedik, mint a víz, de más alkotja, akkor azt, hogy  $H_2O$  molekulák konstituálják, nem tekinthetjük többé a víz fajta-specifikus esszenciális tulajdonságának. Következésképp ki is kell vennünk a helyes szótári definícióból, pontosan úgy, ahogyan számos más tulajdonság esetében eszünkbe se jut megemlíteni a konstitúciót.

Mindez azért tűnhet furcsának, mert az anyagok és természeti fajták túlnyomó többsége esetében a releváns alkotóelemek típusa nem különbözik egy fajtán belül. A „többszörös megvalósulás” legföljebb, ha kivétel. A funkcionálisan azonosított típusok esetében viszont ez a szabály. A kés általában fémből készül, de ha valamilyen speciális üvegből vagy műanyagból készítenénk vágóeszközt, eszünkbe se jutna tagadni, hogy az kés. A természeti fajták esetében viszont empirikus kérdés, mi alkothatja őket. Mindaddig, amíg csak egy mikrofizikai (vagy genetikai) típust találunk, nincs okunk kételkedni, hogy a konstitúció fajta-specifikus esszenciális tulajdonság. De ha kiderül(het)ne – ahogyan azt Kripke, Putnam és mások feltételezik –, hogy a konstitúció radikálisan különbözik (pl. nemcsak izotópok közti eltérésről van szó), akkor az egyben jó indok volna arra is, hogy az adott fajta esetében az illető tulajdonságot természeti, de ne lényegi, esszenciális tulajdonságnak tekintsük.

## X. KÓDA: A RELATIVIZMUS RÉME

Kripke jelentős filozófiai eredményei közé számítjuk, hogy világosan megkülönböztette az analicitás szemantikai, az a prioricitás episztemológiai és a modalitás metafizikai kérdéseit. Az általam javasolt elméletből viszont az következik, hogy analicitás és modalitás nem függetlenek sem egymástól, sem pedig episztemikus állapotunktól. Szó sincs róla azonban, hogy a régi empirista dogmát kívánám védelmembe venni, amely szerint csak az lehet szükségszerű, ami a priori, és csak az lehet a priori, ami analitikus. Mászt állítok: azt, hogy nem szemantikai jellegű tények (metafizikai értelemben vett létezők) befolyásolhatják azt, hogy mi igaz analitikusan és mi nem. Miután pedig a tényeket jórészt empirikus eszközökkel fedezzük fel, episztemikus szituációnk hatást gyakorolhat a jelentésre.

Korábbi fejtegetéseim világos következménye ugyanis, hogy empirikus felfedezéseink befolyásolhatják, mit tekintünk egy természeti fajta szótári értelmében vett esszenciális tulajdonságának.

Ez sokak szerint elfogadhatatlan, mert episztémikus relativizmussal fenyeget. Egy rendes naturalista pedig nem szereti az efféle relativizmust. És ebben én leszek az utolsó, aki vitatkozni fog veled. Csakhogy az általam javasolt értelmezés a természeti fajtákról nem vezet semmiféle relativizmushoz. Alapvető feltételezésem ugyanis, hogy természetről alkotott tudásunk a megismerési folyamattól független tényekre vonatkozik. Amit állítok, és amit sok empirista nem fogadott el, hogy ugyanez igaz fogalmainkra is. Persze nem mindegyikre. De ha igazam van, akkor a természeti fajtákra vonatkozó fogalmaink nem pusztán véletlen vagy konvenció eredményei, hanem maguk is természeti fajták. Ha úgy tetszik, a természet termékei. Minél többet tudunk a természetről, annál inkább.

Mit akarhat ennél többet egy tisztességes filozófiai naturalista?<sup>16</sup>

## IRODALOM

- Altrichter F. 1985/1993. Hit és lehetőség. In uő: *Észérvek az európai filozófiai hagyományban*. Budapest, Atlantisz Könyvkiadó. 329–355.
- Ambrus, G. 2007. *A tudat metafizikája*. Budapest, Gondolat Kiadó.
- Armstrong, D. 1968. *A Materialist Theory of the Mind*. London, Routledge and Kegan Paul.
- Bigelow, J. – B. Ellis – C. Lierse 1992. The World as One of a Kind: Natural Necessity and Laws of Nature. *The British Journal for the Philosophy of Science*. 43. 371–388.
- Boyd, R. 1991. Realism, Anti-Foundationalism and the Ethusianism for Natural Kinds. *Philosophical Studies*. 61. 127–148.
- Carnap, R. 1936–7/1972. Ellenőrizhetőség és jelentés. Ford. Altrichter Ferenc. In Altrichter Ferenc (szerk.) *A Bécsi Kör filozófiája*. Budapest, Gondolat. 377–507.
- Chalmers, D. J. 1996. *The Conscious Mind*. Oxford, Oxford University Press.
- Copi, I. M. 1954. Essence and Accident. *Journal of Philosophy*. 51. 706–719.
- Elder, C. L. 2004. *Real Natures and Familiar Objects*. Cambridge, Mass., The MIT Press.
- Ellis, B. – C. Lierse 1994. Dispositional Essentialism. *Australasian Journal of Philosophy*. 72. 27–45.
- Fodor, J. A. 1974. Special Sciences, or the Disunity of Sciences as a Working Hypothesis. *Synthese*. 28. 97–115.
- Goodman, N. 1955. *Fact, Fiction and Forecast*. Cambridge, Mass., Harvard University Press.
- Gorman, M. 2005. The Essential and the Accidental. *Ratio*, 18. 276–289.
- Hume, D. 1748/1973. *Tanulmány az emberi értelemről*. Ford. Vámosi Pál. Budapest, Magyar Helikon.
- Huoranszki Ferenc 2001. *Modern Metafizika*. Budapest, Osiris.
- Huoranszki Ferenc 2008. Tudatosság és intencionalitás. *Világosság*. 49/11–12. 133–142.
- Jackson, F. 1998. *From Metaphysics to Ethics. A Defence of Conceptual Analysis*. Oxford, Clarendon Press.

<sup>16</sup> Köszönettel tartozom Ambrus Gergelynek és a tanulmány két anonim referensének egy korábbi változathoz fűzött fontos kommentárjaikért.

- Johnston, M. 1992. How to Speak of the Colors. *Philosophical Studies*, 68, 221–263.
- Kripke, S. 1972/2007. *Megnevezés és szükségszerűség*. Ford. Bárány Tibor. Budapest, Akadémiai Kiadó.
- LaPorte, J. 1996. Chemical Kind Term Reference and the Discovery of Essence. *Noûs*. 30. 112–132.
- Lewis, D. 1986. *On the Plurality of Worlds*. Oxford, Basil Blackwell.
- Locke, J. 1690/2003. *Értekezés az emberi értelemről*. Ford. Vassányi Miklós és Csordás Dávid. Budapest, Osiris.
- Lowe, E. J. 2006. *The Four-Category Ontology. A Metaphysical Foundation for Natural Science*. Oxford, Clarendon Press.
- Mellor, D. H. 1977. Natural Kinds. *British Journal for the Philosophy of Science*. 28. 299–331.
- Owens, D. 1992. *Causes and Coincidences*. Cambridge, Cambridge University Press.
- Prior E. – Pargetter, R. – Jackson, F. 1982. Three Theses about Dispositions. *American Philosophical Quarterly*. 19. 251–256.
- Putman, H. 1975. The Meaning of „Meaning”. *Minnesota Studies in the Philosophy of Science*. 7. 215–271.
- Quine, W. V. 1960. *Word and Object*. Cambridge, Mass., Harvard University Press.
- Quine, W. V. 1969/1999. Természeti fajták. Ford. Kampis György. In Forrai G. – Szegedi P. (szerk.) *Tudományfilozófia*. Budapest, Áron Kiadó. 383–398.
- Quine, W. V. 1953/2002. Referencia és modalitás. Ford. Eszes Boldizsár. In uő: *A tapasztalattól a tudományig*. Szerk. Forrai Gábor. Budapest, Osiris Kiadó. 225–251.
- Salmon, N. 1979. How *Not* to Derive Essentialism from the Theory of Reference. *Journal of Philosophy*. 76. 703–25.
- Shoemaker, S. 1980. Causality and Properties. In P. van Inwagen (szerk.) *Time and Cause: Essays Presented to Richard Taylor*. Dordrecht, Reidel. 109–135.
- Wiggins, D. S. 2001. *Sameness and Substance Renewed*. Cambridge, Cambridge University Press.

## Fajtáink eredete\*

A természeti fajták metafizikájának kérdése talán egyetlen tudományterületen sem váltott ki olyan heves vitákat, mint a biológia esetében. A biológiai fajok léte és jellegzetességeik meghatározására vonatkozóan számos eltérő álláspont alakult ki, egyvalami azonban rendkívüli fontosságú a naturalista tudományfelfogás számára: a projektnek számot kell adnia a társadalmi jelenségek biológiaiakra történő redukálásáról, amennyiben azokat a kémiai, majd a fizikai jelenségek szintjére szeretné visszavezetni, így biztosította a tudományos elméletek magyarázóerejét. A biológiai fajokkal kapcsolatos esszencializmus ígéretes megoldási javaslatot kínált a problémára: az elmélet egyrészt redukcionista, mivel a természeti fajtákon kívüli esetleges egyéb (történeti, társadalmi, kulturális) fajtákat az előbbiekre redukálhatónak tartja, másrészt pedig a fizikának, kémiának és biológiának kitüntetett szerepet szánva a többi tudományt azok módszereivel művelhetőnek mutatja be.<sup>1</sup> A redukció végeredményeképp a fizika válik a természeti és az arra redukált társadalmi jelenségeket egyaránt magyarázni képes tudománnyá – az általa alkalmazott matematikai apparátus segítségével.

Amennyiben tehát a *homo sapiens* is pusztán biológiai kategória, úgy a tudomány feladata az esszencia meghatározása után az annak köszönhetően megvalósuló felszíni tulajdonságok, azaz az esetleges, akcidentális jellegzetességek magyarázata. A biológiai esszencializmust azonban számos kritika érte, s az alternatívák sokfélesége arra enged következtetni, hogy a faj fogalma, noha alapvető kategóriája a biológia tudományának, máig sem szolgál olyan szilárdan meghatározott kiindulási pontként, mint pl. a kémiai elemek csoportja.

Mi alapján sorolunk bizonyos egyedeket ugyanazon fajba tartozónak? Mi tekinthető érvényes, használható klasszifikációnak? Az alábbiakban megkísérlem bemutatni az esszencializmus-kritikák alternatív javaslatait, majd emellett érvelek, hogy a különféle biológiai klasszifikációelméletek, bármennyire „megengedők” és radikálisak legyenek is a fajok felosztásának kritériumaival kapcsol-

\* A tanulmány írása során számos beszélgetést folytattam Demeter Tamással – tanácsaiért és kritikáiért rendkívül hálás vagyok.

<sup>1</sup> Wilkerson 1998. 225.



latban, nem képesek áthidalni a társadalmi és biológiai fajták közötti fogalmi, minőségi különbségeket, ennek kapcsán pedig a redukcionista kép létjogosultságát sem tudják meggyőzően alátámasztani.

## I. A TRADICIONÁLIS KÉP HIBÁI

A természeti fajtákkal kapcsolatos esszencializmus Hilary Putnam és Saul Kripke alapvető munkái<sup>2</sup> óta jelentős szerepet játszik a természettudományos fogalomalkotás, az induktív általánosítás és előrejelzés filozófiai megalapozásában. Az eredetileg arisztotelianus gyökerekkel bíró esszencialista felfogást új megközelítésben felvető tudományfilozófiát az elmúlt évtizedek során számos – mind filozófiai, mind természettudományos elméleteken alapuló – bírálat érte, melyek igyekeztek kimutatni az esszencialista metafizika tarthatatlanságát.

Az alábbiakban az esszencializmusnak ezzel az 1960-as, 1970-es években új megalapozást nyert irányzatával foglalkozom, amely szerint a természetben fellelhető természeti fajták szolgáltatják az alapot az objektív természettudományok számára, ezek segítségével válik lehetővé az adott fajtán belüli általánosítás, magyarázat és predikció anélkül, hogy fajtáink összes példányát meg kellene vizsgálnunk a tudományos eredmények eléréséhez. Az ugyanazon fajtába tartozó individuumok hasonlóságát pedig valamilyen mikrostrukturális sajátosság, valamilyen esszencia biztosítja: az adott fajhoz tartozás szükséges és elégséges feltétele, hogy a vizsgált példány vagy egyed rendelkezzen azzal a szabad szemmel nem látható tulajdonsággal, amely alapján egyértelműen besorolható a megfelelő kategóriába. Így lehetséges ugyanis elvonatkoztatni a felszíni jellemzők esetleges eltéréseitől: ami a példány *lényegét* alkotja, az a makroszint alatt keresendő. Kémiai elemek esetében ilyen például az elemszám, élőlényeknél a meghatározott genetikai kód. A vizsgált entitások olyan tulajdonságai ezek, amelyek hiányában nem tarthatnának az adott fajtába, meglétük tehát szükségszerű minden egyes példány esetében – szemben a felszíni jellegzetességekkel, melyek pusztán akcidentálisak (vagy a putnami képet követve: a nominális esszenciát alkotják).<sup>3</sup> A tudomány feladata ezen esszenciák felfedezése, s segítségükkel a természet jelenségeinek magyarázata, illetve előrejelzése. A tradicionális esszencialista kép tehát realista: a világ jelenségei fajtákba rendeződnek, amelyekről igaz ismereteket szerezhethünk a természettudományok segítségével.

A tradicionális képet ért bírálatok számára a „valódi esszenciák” megléte jelenti a fő támadási felületet: ha vannak ilyenek, akkor a tudomány képviselőinek már fel kellett vagy hamarosan fel kell fedezniük őket – ami az esszencializmus egyik állandó példajaként szolgáló biológiai fajták esetében leküzdhetetlen-

<sup>2</sup> Putnam 1975, illetve Kripke 2007.

<sup>3</sup> Putnam 1975. 268–271.

nek tűnő nehézségekbe ütközik. A „faj” (*species*) fogalma a biológusok körében sem bír általánosan elfogadott jelentéssel – Reydon (2006) például *négy* különböző fajfogalomról ír, de ennél jóval nagyobb számokkal is találkozhatunk.<sup>4</sup> Az esszencialista képtől viszont azt várnánk, hogy ne engedje meg a lehetséges fajfogalmak ilyen nagy számát – az esszencia megléte az adott fajba tartozás szükséges és elégséges feltétele kellene legyen. A mikrostrukturális esszencia legesélyesebb jelöltjének, az adott faj DNS-mintájának feltérképezése és a genomok felfedezése viszont épp ellenkező eredménnyel szolgált. Ahelyett, hogy a biológusok számára rendelkezésre állt volna a fajhoz tartozást egyértelműen meghatározó sajátosság, az adott fajokon belüli egyedek genomjai is számos eltérést mutattak egymáshoz képest<sup>5</sup> – amennyiben tehát a genom lenne a valódi esszencia, oly irreális mértékben sokszorozódna meg a biológiai fajok száma, hogy a tudományterület törvényei igen kis magyarázóerővel bírnának, meglehetősen szűkös területet (egy-egy egyed) lefedve.

Az evolúciós fejlődés további akadályokat állít az esszencializmuson alapuló biológiai klasszifikáció elé. Az esszenciális tulajdonság ugyanis olyan jellegzetesség, amellyel az adott faj minden egyedének rendelkeznie kell, ugyanakkor még a fajhoz a törzsfajlódási fán (az „élet fáján”) legközelebb álló más fajok egyedei sem birtokolhatják. Az evolúció azonban számos példával szolgál arra, hogy ilyen tulajdonságot lehetetlen felfedezni – még akkor is, ha az esszenciát olyan tulajdonságok csoportjaként definiáljuk, melyek külön-külön szükségszerűek és együttesen elégségesek lennének.<sup>6</sup> Az egymással szoros kapcsolatban lévő, ugyanakkor más-más fajba tartozó élőlények esetében gyakran fejlődnek ki rendkívül hasonló jellegzetességek, melyek a közös ősüktől származnak, és az evolúciós közelség révén nem térnek el jelentős mértékben egymástól – ugyanakkor egy adott fajon belül is megjelenhetnek olyan mutációk, melyek megakadályozzák, hogy az „esszenciális” tulajdonság a faj minden egyedében felfedezhető legyen.<sup>7</sup> Mindezek a problémák azonban csak az esszencializmus biológiai alkalmazhatóságát vonják kétségbe, a fajok elhatárolásának kérdésében nem szolgálnak egyértelmű útmutatókként. Az esszencializmus kiküszöbölése nem nyújt egyértelmű megoldást arra nézve, hogy mi alapján tagolódnak a természet fajokra, így a taxonómusok problémái nem oldódnak meg a lényegi tulajdonságok elméletének elvetésével.

Hull 1965 szerint az evolúció sajátosság dilemma felvetésére készíti a biológiai taxonómia művelőit: „Ha [a taxonómusok] elfogadják az evolúció elméletét mint a természetes klasszifikációt megalapozó elvet, úgy fel kell adniuk annak reményét, hogy valaha is sikerül valódi fajokra lelniük. Ha a valódi fajok létét

<sup>4</sup> Lásd Ereshefsky 2008, Kitcher 1984.

<sup>5</sup> Ereshefsky 2008. 99.

<sup>6</sup> Hull 1965 példája a fél-gerinchúrosokkal kapcsolatban ugyan a fajnál magasabb klasszifikációs szintekre vonatkozik, de az elképzelés kiterjeszthető a fajok elhatárolására is – lásd Boyd 1991, 1999. fajmeghatározását, melyről a későbbiekben lesz szó.

<sup>7</sup> Ereshefsky 2008. 101–102.

szeretnék megőrizni, akkor viszont a természetes klasszifikációról kell lemondaniuk.”<sup>8</sup> Az evolúció elmélete alapján a fajok oly mértékben változnak, hogy nincs értelme „valódi fajokról” beszélni – attól függően persze, hogy milyen fajfogalommal dolgozunk. Hull dilemmája a monista fajmeghatározásokat érinti érzékenyen: azokat, amelyek feltételezik, hogy a biológiai fajok elhatárolásának egy és csak egy lehetséges módja van. A legbefolyásosabb monista elméletek a fajokat reprodukív izolációban élő (azaz egészséges utód nemzésére képes) egyedek (biológiai fajfogalom); közös őshöz visszavezethető egyedek (filogenetikus fajfogalom); vagy meghatározott ökológiai feltételekhez alkalmazkodó egyedek (ökológikus fajfogalom) összességéként fogják fel. A „valódi fajok” pedig azért lehetnek problematikusak bármelyik elmélet szempontjából, mert az evolúciós dinamika eredményeként előfordulhat, hogy egy adott fajban megjelenő mutáció más fajokkal történő kereszteződést vagy más feltételekhez való alkalmazkodást tesz lehetővé.<sup>9</sup> Amennyiben tehát a monista elméletek szeretnék megtartani a faj mint természeti fajta fogalmát, a valódi esszenciáknál megengedőbb klasszifikációs bázist kell találniuk – olyat, amely egyrészt számot vet az evolúciós elmélet következményeivel, másrészt elég stabil ahhoz, hogy a biológia mint tudomány magyarázóerejét fenntartsa.

Richard Boyd (1991) azokra a fajspecifikus tulajdonságokra helyezi a hangsúlyt, amelyek oksági erejüknel fogva szerepet játszanak az empirikus általánosítás, magyarázat és predikció területén – anélkül, hogy ezen tulajdonságok meglétét az adott faj összes példánya esetében szükségszerűnek vagy elégségesnek tartaná: „A biológiai indukcióra és magyarázatra az tesz alkalmassá minden egyes biológiai fajt, hogy annak egyedei rendelkeznek bizonyos, tökéletlenül elosztott és egymással homeosztatikus kapcsolatban álló morfológiai, fiziológiai és viselkedésbeli jellegzetességekkel.”<sup>10</sup> Ezek a tulajdonságok és a hozzájuk kapcsolódó oksági mechanizmusok biztosítják, hogy a fajon belül fennmaradjon a homeosztázis; a fajok Boyd szerint „homeosztatikus tulajdonságnyaláb”-fajták. A klasszifikációnak pedig figyelembe kell vennie az evolúcióból következő indetermináltságot: egy új fajta kialakulása minden esetben függ a megelőző és az új között megjelenő közvetítő populáció meglététől – aminek figyelmen kívül hagyása összeegyeztethetetlen az öröklődő fenotipikus variációk elvével.<sup>11</sup> A fajok tehát szoros kapcsolatban vannak elődjeikkel – az evolúciós változások és környezeti hatások egyaránt formálják a homeosztázisuk fenntartásában létfontosságú tulajdonságokat, s ez alapján joggal nevezi őket Millikan (1999) történeti fajtáknak.<sup>12</sup> A meghatározó sajátosságok tehát ebben a képben már nem feltét-

<sup>8</sup> Hull 1965. 319.

<sup>9</sup> Természetesen a fent említett fajfogalmak számos egyéb problémát is felvetnek: a biológiai fajfogalom például nem tud számot adni az aszexuális fajokról (Kitcher 1984).

<sup>10</sup> Boyd 1991. 142.

<sup>11</sup> U.o.

<sup>12</sup> Mellyel Boyd 1999 is egyetért.

lenül az adott faj egyedeinek mikroszerkezetében keresendők – a homeosztázis fenntartásában extrinzikus elemek is fontos szerepet játszanak.

Noha a homeosztatikus tulajdonságnyalábok elmélete sem kínál megoldást minden, fajokkal mint természeti fajtákkal kapcsolatos problémára (nem szolgál magyarázattal a fajokon belüli szignifikáns eltérésekkel kapcsolatban, valamint nem szán kitüntetett szerepet a genealógiának – egyéb faktorok is közrejátszhatnak a hasonlóságok kialakításában),<sup>13</sup> az mindenképp elmondható róla, hogy a hagyományos esszencializmusnál *megengedőbb*, azaz nem köti a természeti fajták kialakulását és esetleges változását pusztán intrinzikus jellegzetességeikhez: a klasszifikáció során biológiai és történeti megfontolásoknak egyaránt plauzibilitást garantál. Millikan (1999) szerint épp ez a történeti háttér biztosítja, hogy általánosításaink megbízhatóak legyenek: ha van valamilyen lényegi sajátosság, ami a fajok egyedeit egymáshoz köti, az a közös történelmükből fakadó kapcsolat – amely nem vezet ugyan kivétel nélküli természeti törvényekhez, de alapot ad nem véletlenszerű általánosításokra.<sup>14</sup> A faj tehát ez esetben nem lesz a hulli értelemben vett „valódi faj”, mivel egyedei nem változatlan tulajdonságokkal bíró példányok. A klasszifikáció maga sem lesz azonban szükségszerűen azon rendszer leírása, amely alapján a természet fajtákra oszlik, hiszen a természeti hatások önmagukban nem elégséges feltételei a homeosztázis fenntartásának.

A boydi koncepció tehát igyekszik megőrizni a természeti fajtákat a biológiai fajok metafizikai háttereként – továbbra is fennmarad azonban a monizmus problémája: noha a tulajdonságnyalábok elmélete az esszencializmusénál szélesebb spektrummal közelít a fajok kérdéséhez, háttér-feltételezésként megmarad az egyetlen valódi klasszifikációs módszer megléte. Mint azt fentebb már említettem, az „egyetlen módszer” tekintetében a biológusok véleménye is megoszlik – a különböző al-tudományterületeken belül dolgozók más-más módszert tekintenek kitüntetett fontosságúnak. A megosztottságra adott válaszok közül a *fajpluralizmus* az eddigieknél is nagyobb teret enged a klasszifikáció művelőinek – közelítve ezzel a természeti és társadalmi fajták ontológiai hátterét. A továbbiakban a pluralista nézetek (Kitcher és Dupré) rövid ismertetése után megvizsgálom, hogy az általuk posztulált fajfogalmak képessé tehetik-e a naturalizmus támogatóit egy egységes tudományos módszertan sikeres védelmezésére.

## II. ELMOSÓDÓ HATÁROK: PLURALIZMUS A TERMÉSZETBEN

A szintén pluralista koncepciót képviselő Ereshefsky (2008) felveti a kérdést: hogyan klasszifikálunk három, azonos helyen élő rovarpopulációt (*A*, *B* és *C*), melyek közül kettő (*A* és *B*) képes közös utódot létrehozni, egy viszont aszexual-

<sup>13</sup> Ereshefsky 2008. 103–104.

<sup>14</sup> Millikan 1999. 54–55.

lisan szaporodik; kettő (*B* és *C*) közös ökológiai szereppel bír – továbbá mindhárom külön-külön visszavezethető meghatározott ősokhöz. A biológiai fajfogalom alkalmazásával egyetlen fajunk lesz – az egymással szaporodni képes populációk tartoznak ide ( $A + B$ ), az aszexuális viszont ebben a klasszifikációban értelmezhetetlen. A filogenetikus fajfogalom három fajt eredményez az ősökre való hivatkozással (*A*, *B*, *C*), az ökológiai fajfogalom pedig kettőt: az azonos szereppel bírók tartoznak azonos fajhoz (*A*, illetve  $B + C$ ).<sup>15</sup> A dilemmára két válasz adható: feltétezhetjük, hogy létezik egyetlen érvényes klasszifikációs eljárás, és a biológia mai fejlettségi szintje még nem elégséges ahhoz, hogy rátaláljunk (amely feltételezés erősen hasonlít a valódi esszenciák tanáéhoz, melyek felfedezését szintén a jövő nemzedékektől várták) – vagy azt mondhatjuk, hogy a természet nem rajzol univerzális határvonalakat, a többféle klasszifikáció egyaránt érvényes lehet bizonyos körülmények között. Ha azonban így van, mi szabja meg, hogy melyik definíciót használjuk – mi alapján szegmentálható a valóság úgy, hogy a biológiai általánosítások és predikciók érvényessége és megbízhatósága megmaradjon?

Kitcher (1984) szerint a fajok nem feltétlenül természeti fajták (bár mutatnak némi hasonlóságot azokkal), hanem organizmusok halmazai. A halmazok viszont nem pusztán egyféle rendezési elv alapján különíthetők el egymástól, nem tartanak igényt kizárólagosságra egyetlen, kitüntetett tulajdonság vagy tulajdonságok alapján felállított klasszifikáció sem.

Számos [biológiai] reláció létezik, melyet felhasználhatunk a fajok csoportosításához. Nincs azonban egyetlen olyan kivételes reláció sem, amely abból a szempontból privilegizált helyzetben lenne, hogy az általa létrehozott csoportok minden biológus kérdésére választ adnának és alkalmazhatók lennének az organizmusok összességére. Röviden, a faj kategóriája heterogén.<sup>16</sup>

Kitcher tehát nem vitatja, hogy a biológia képes a fajok elhatárolására morfológiai sajátosságok, filogenetikus szempontok vagy jellemző tulajdonságok (pl. szaporodási képesség) alapján, pusztán azt vonja kétségbe, hogy bármelyik is a biológiai valóság egyetlen kulcsát adná alkalmazói kezébe.

A biológiai magyarázatok ugyanis Kitcher szerint maguk sem lehetnek homogének és univerzálisan alkalmazhatók mindenfajta vizsgálódás esetében. Különböző érdekek motiválhatnak a biológia területén belül kutató különböző tudósokat, s az általuk felállított klasszifikációk ezt a speciális szempontot tükrözik – magyarázataik pedig ennek következtében az adott részterületen belül bírnak jelentőséggel. Egy bizonyos organizmus tulajdonságait kívánhatjuk vizsgálni azért, hogy felderítsük, pontosan milyen folyamatok hozzák azokat létre

<sup>15</sup> Ereshefsky 2008. 105.

<sup>16</sup> Kitcher 1984. 309.

az adott szervezetben, és milyen szerepet játszanak az élőlény túlélésének elősegítésében – de vizsgálhatjuk azért is, hogy megtudjuk, milyen evolúciós folyamatok és szelekciós mechanizmusok játszottak szerepet abban, hogy azok egyáltalán kialakuljanak. Az előbbi vizsgálódáskor *strukturális*, az utóbbi esetben pedig *történeti* magyarázatot keresünk – és egyikről való tudomásunk sem helyettesítheti a másikat.<sup>17</sup>

Kitcher felvetéseiben tovább finomodik a fajok elhatárolásának elméleti háttere: a pluralizmus nem csupán az esszenciális tulajdonságokról mond le, hanem a természeti jelenségek „természetes” felosztásáról is – a lehetséges fajfogalmak a tudományos megismerési érdekek alapján konstruálódnak, nem céljuk többé a természetben meglévő univerzális rend feltérképezése, mivel magát az univerzális rend létjogosultságát kérdőjelezi meg. A rend többféleképpen hozható létre, és az adott területen belül magyarázóerővel bíró koncepciók egyike sem alárendelt az összes többinek, amíg tudományos magyarázatok elérésére törekszik. A természeti jelenségek tehát több klasszifikáció alapján is *megismerhetők*, a legitimitást a tudományos igény biztosítja.

A fajokkal kapcsolatos pluralizmus radikálisabb elmélete a Dupré (1981) által képviselt „promiszkuózus realizmus”, amelyben még a Kitcher által támasztott igények (a kutatók megismerési érdekei) sem tarthatnak számot kizárólagos érvényességre. A fajpluralizmus egyéb képviselőitől Dupré felfogása abban tér el leginkább, hogy a klasszifikációk sokfélesége nála nem pusztán a tudományon belüli, más-más jellegzetesség hangsúlyozásával megalapozott felosztások legitimitását garantálja, hanem a hétköznapi nyelvhasználók, a „népi biológia” eltérő besorolásaiét is. A magasabb taxonómiai szintek Dupré szerint minden esetben mesterségesen létrehozott kategóriák, egyedül a faj fogalma lenne alkalmas arra, hogy a természetben rejlő tényleges elkülönüléseket hitelesen tükrözze – ám a fajok esetében is olyan mértékű heterogenitással vagyunk kénytelenek dolgozni, amely nem pusztán egyetlen klasszifikációs módszert tesz lehetővé.

Dupré – a putnami és kripkei intuíciókból kiindulva – a fajok és magasabb klasszifikációs szintek neveinek tudományos és hétköznapi használatát összehasonlítva arra jut, hogy a kettő között számos esetben figyelhető meg eltérés, amikor pedig a kettő egybevágh, azt többnyire funkcionális/pragmatista szempontok motiválják. A magasabb klasszifikációs szintek esetében szembetűnőbb a különbség: a tudományon kívüliek számára kevésbé jelentős törzsek, rendek és osztályok megkülönböztetése a hétköznapi nyelvhasználat során nagymértékben eltérhet a tudományostól (pl. bogarak és rovarok, nyitva- és zárvatermők), a fajok csoportosítása pedig olyan antropocentrikus megkülönböztetési jegyeken (pl. méret, íz, felhasználhatóság) alapulhat, amelyek a tudományos megalapozást teljességgel nélkülözik.<sup>18</sup> Dupré szerint azonban ezek a klasszifi-

<sup>17</sup> Kitcher 1984. 320–321.

<sup>18</sup> Dupré 1981. 80–81.

kációk sem rosszabbak a tudományos besorolásoknál – pusztán más szempontok alapján osztják fel a természet jelenségeit; ezért nevezhető az elmélet „promiszkuoózus realizmusnak”.

A realizmus abból a tényből következik, hogy számos olyan hasonlósági relációt találhatunk, amelyek alapján a különféle érdekeinknek megfelelően elkülöníthetjük az organizmusok különböző osztályait; a promiszkuoóztatás pedig abból, hogy ezek egyike sem privilegizált.<sup>19</sup>

Olyannyira nem privilegizált, hogy a tudományos rendszertannak ellentmondó kategóriák sem számítanak lenézendőnek vagy alacsonyabb rendűnek – mivel a taxonómia maga sem képes felmutatni általánosan érvényes rendezési elvet. A természeti jelenségek a leghétköznapibb igényeket szem előtt tartva is *szoportosíthatók*, sőt, bizonyos körülmények között felmerülhet az a kérdés is, hogy nem volna-e célszerűbb adott esetben a tudományon kívüli használatot beemelni a tudományos praxisba.

Dupré tehát a kitcheri engedékenységnél is szélesebbre tárja a kaput: a szükségszerű esszenciális tulajdonságok helyét (némi túlzással) *bármí* átveheti. A biológiai kategóriák még ezen legmegengedőbb felfogása sem alapozhatja meg azonban a tudományok módszertani egységét, a társadalmitól a biológiaiig vezető úton ugyanis olyasmi történik fogalmainkkal, amelyre a naturalista megközelítés nem tud kielégítő magyarázatot adni.

### III. KONCEPTUÁLIS GYÖKEREK TERMÉSZETBEN ÉS TÁRSADALOMBAN

A kék bálna az emlősök osztályába tartozó állatfaj, melynek emlős volta mindenki számára egyértelmű, aki tanulmányai során megismerkedett biológiai definíciókkal is. Könnyen elgondolható azonban egy olyan civilizáció, amelynek tagjai nem az elevenszülés, az utódok táplálása vagy éppen a vérmérséklet alapján húzzák meg klasszifikációik alapvonalait, hanem egyéb, számukra fontos tulajdonságokat vesznek lajstromba az adott állatfajok besorolásánál. Esetükben például a kék bálnára leginkább jellemző, hogy vízben él, és húsa emberi fogyasztásra alkalmas. Ebben viszont nem különbözik a halak legtöbbszörétől, így amennyiben ezek alapján sorolnánk be bálnáinkat valamilyen biológiai osztályba, úgy kétségkívül értelmes felvetés lenne hálnak tekintenünk őket. Dupré (1999) ezt az álláspontot kívánja bemutatni és megvédeni, promiszkuoózus realista koncepcióját alátámasztandó.

A bálnákról alkotott *képünk* szempontjából konstitutív, hogy mit tartalmaz bálna-fogalmunk: mik a fontos és mik a kevésbé fontos tulajdonságai. A nyugati

<sup>19</sup> Uo. 82.

civilizáció oktatási rendszereiben visszatükröződő természettudományos osztályozás kategóriái szerint legfontosabb jellemzői a melegvérűség, az elevenszülés – ezek alapján pedig az emlősök osztályába tartozás.<sup>20</sup> Noha a biológia nem tekintené érvényes besorolásnak, számos kultúrkörben élhet olyan, a bálnákat és halakat egy osztályba klasszifikáló felfogás, amely alapján a hétköznapi nyelvhasználók eligazodnak a természeti jelenségek világában.

A bálnákat magukat ez azonban egyáltalán nem befolyásolja: a kék bálnát tekinthetjük a többi állatfajtól esszenciális tulajdonság tekintetében, homeosztatisz tulajdonságnyalábok vonatkozásában vagy pusztán morfológiai sajátosságok alapján eltérőnek, azt egyik esetben sem kérdőjelezzük meg, hogy klasszifikációnk célja a *felismerésük* és egyéb fajoktól való *elkülönítésük*. Amennyiben a nyugati tudományosságtól teljesen különböző civilizációk egy adott faj egyedeit más-képp, számunkra tudománytalanul határozzák meg, az az egyedek szempontjából irreleváns – kiderülhet, hogy tévúton járnak, de az is előfordulhat, hogy a tudomány aktuális fejlettségi szintje számunkra is lehetetlenné teszi a megcáfolatlan kategória-határok megrajzolását. Bálnák és egyéb tengeri élőlények például közös platformra kerülhetnek pusztán élőhelyük miatt és azért, mert csillapítják annak éhségét, aki kifogja és elfogyasztja őket (sőt, Dupré azt a lehetőséget is felveti, hogy a bálnák kivonása a halak közül végbemehetett morális alapokon is: mivel általában intelligensebb állatoknak tartjuk őket egyéb halaknál, elfogyasztásukat nagyobb valószínűséggel gondoljuk elítélendőnek, mint a kevésbé értékesnek tartott vízi élőlényekét).<sup>21</sup>

Elgondolható tehát az állatok olyan felosztása, amely a bálnákat minden, számunkra biológiailag releváns tulajdonságuktól függetlenül csoportosítja – egy- valami azonban kevésbé tűnik kivitelezhetőnek: az, hogy pusztán a rájuk való hivatkozással *alkossuk meg* a bálnákat a maguk fizikai és biológiai valójában. Bálnafogalmunk számos változáson mehet keresztül: érvényét vesztheti, és új, még tudományosabb meghatározással helyettesíthető, de megmaradhat hétköznapi érdekeinket tükröző, praktikus szempontokat figyelembe vevő konstruktumnak is (esetleg a tudományos és mindennapi meghatározás együttélése alapján „mindkét urat is szolgálhatja”). Ha a fajok elhatárolásának kérdése évszázados problémát jelent is a biológia művelői számára, azt egyik felfogás képviselői sem állítják, hogy magát az elhatárolandót is ők hoznák létre a róla való fejtegetéseik során.

Wilkerson (1998) az esszencializmusról szólván említi a tudományok módszertani egységét: a fizika, kémia és biológia azért lehetnek privilegizált helyzetben, mert ők hivatottak a dolgok valódi természetéről számot adni; minden egyéb

<sup>20</sup> Természetesen a kurrens biológiai klasszifikációkban a felsorolt tulajdonságok egyike sem képezi a bálnák „valódi esszenciáját”, minthogy annak megeléje – mint az fentebb látható volt – immár nem tartozik a biológia fő motivációi közé.

<sup>21</sup> Dupré 1981. 468.



tudományág ezeket az eredményeket felhasználva lehet csak képes érvényes és megbízható magyarázatok és predikciók megfogalmazására, a megfelelő redukciók végrehajtása után.<sup>22</sup> Mint azt fentebb láttuk, a biológiai fajokkal kapcsolatos esszencializmust számos egyéb elméleti megfontolás megdönteni látszik – a természettudományos (matematizáló) módszer univerzális használhatóságának kérdését ez azonban nem befolyásolja. A természettudományok segítségével megtudhatjuk, hogy milyen természetű dolgok léteznek a világban, az ember pedig maga is ezek közé a dolgok közé tartozik. Így amennyiben a *homo sapiens* biológiai fajt alkot (akár felfedezhető esszenciális tulajdonságok hiányában is), elgondolható az emberi (társadalmi) jelenségek biológiai alapokra redukált magyarázata, ami pedig tovább redukálható kémiai és végül fizikai terminusokra. Ebbe a redukcionista képbe azonban nem fér bele a természeti és társadalmi jelenségek eltérő mértékű fogalmi meghatározottsága, ami a matematikai apparátus univerzális használhatóságát hirdető tudományos módszerek létevel kapcsolatban komoly ellenvetések megfogalmazására adhat alapot.<sup>23</sup>

Láthattuk, bálnafogalmunk szempontjából meghatározó, hogy milyen kép alakult ki kultúránkban a bálnáról. Társadalmi jelenségeinkkel kapcsolatban azonban egy ennél erősebb állítást is tehetünk: az ő esetükben a *létük* szempontjából konstitutív, hogy mit involválnak fogalmaink. Ha megszűnnénk használni őket, a jelenségek maguk szűnnének meg, míg ha hirtelen elfelejtenénk a bálna fogalmát, és soha nem használnánk azt többé, a világ bálnái nem pusztulnának el ennek következtében.

Hacking (1986) a társadalmi jelenségek vizsgálatára bevezetett kategóriáinkat épp ezért nem besoroláznak, hanem „feltalálásnak” nevezi: nem olyasvalamit próbálunk meg megszerezni, ami tőlünk függetlenül már a rendszertanunk bevezetése előtt is létezett – a kategorizációink *maguk* válnak a jelenségek konstitutív alapjává, nem meg-, hanem feltalálják őket. További, természeti jelenségektől lényegesen különböző tulajdonságuk, hogy amíg azok az alkalmazott kategorizációkra való tekintettel teljességgel közömbösek, addig a társadalmi fajták egyedei a legtöbb esetben tudatában vannak a besorolásuknak – képesek reagálni rá, megváltoztatni viselkedésüket vagy bizonyos esetben a magát a kategóriát is.<sup>24</sup> Mindezt pedig azért tehetik meg, mert a társadalmi kategóriák

<sup>22</sup> Wilkerson 1998. 225–226.

<sup>23</sup> Az „univerzális módszer” kifejezést a természet- és társadalomtudományos módszerek különbségével kapcsolatban alkalmazva nem szeretném azt a látszatot kelteni, hogy valamilyen *egyetlen* természettudományos módszer tekintetében a különböző tudományágak között teljes egyetértés lenne. Jelen tanulmány szempontjából a természet- és társadalomtudományok közötti eltérés bír elsődleges fontossággal: az, hogy a matematikai alapú módszerek alkalmasak a természeti jelenségek vizsgálatára, a társadalmiak esetében azonban a jelenségek eltérő minősége kapcsán korlátokba ütköznek. A természettudományokon belüli megosztottság tekintetében lásd Suppes 1976, illetve Hacking 1996.

<sup>24</sup> Hacking 1999 ennek kapcsán nevezi az emberi fajtákat interaktívoknak, a természetieket pedig indifferensnek.

által létrehozott jelenségek fogalmi konstruktumok, szemben például a biológiai fajokkal, amik – bármilyen felosztást tekintünk is érvényesnek – sohasem a kategóriáknak köszönhetik létüket. A különbségek jól érzékeltehetőek a fent elemzett biológiai fajokat a házasság intézményével és az ahhoz kapcsolódó társadalmi fajtákkal (férj, feleség) összehasonlítva.

A házasság fogalma kultúránként eltérő jelentéstartalmat involválhat, bizonyos civilizációk köthetik megvalósulását jogi aktusokhoz, mások hivatkozhatnak vallásos alapokra, megint mások akár azt is vallhatják, hogy két ember huzamosabb ideig fennálló együttélése valósítja azt meg. Képzelnünk azonban el egy ennél is radikálisabb változatot: egy kultúrát, melyben a házasság fogalma teljességgel ismeretlen. Ebben a képzeletbeli kultúrában is élnek emberek közös otthonokban, azonban semmi sem garantálja, hogy ezt az állapotot hosszabb távon is fenntartsák: szabadon lakhatnak együtt akárkivel, szabadon hozhatnak létre utódokat, tulajdonaik megosztása pedig teljes mértékben saját belátásukra van bízva. Elképzelt kultúránkban nem csupán férjek és feleségek különféle tulajdonságairól (jogok, kötelezettségek) értelmetlen beszélni, hanem magukról e kategóriákról is. Amit mi – tévesen – férjként és feleségként azonosítanánk, az számukra ugyanis *nem létezik*. Amennyiben pedig megismertetnénk őket a házasság fogalmával, kétségtelenül megváltozna az önmagukról alkotott képük: olyan új szempontból lennének képesek reflektálni saját magukra, múltjukra és cselekvési lehetőségeikre, amely korábban elképzelhetetlen lett volna számukra.

Saját civilizációinkban is megjelenhetnek olyan társadalmi fajták, amelyekről nem lehetett fogalmunk megalkotásuk előtt – Hackinggel szólva megemlíthető példaként a „gyermekbántalmazás” vagy a „többszörösen összetett személyiség” fogalmának bevezetése és mindezek kulturális és társadalomtudományos változásokat előidéző következményei.<sup>25</sup> A 20. század közepe előtt ezek a fogalmak ismeretlenek voltak, ami egyúttal azt is jelenti, hogy olyan emberek sem léteztek, akiket ma gyermekbántalmazónak vagy többszörösen összetett személyiséggel rendelkezőnek hívunk. Természetesen ez nem jelenti azt, hogy a korábbi időkben ne követtek volna el kegyetlenségeket gyermekek kárára – azonban a fogalom használata előtt ezek az atrocitások semmiben sem különböztek egyéb erőszakos cselekményektől. A morális megítélés megváltozása, az elkövetők új emberi fajtába történő besorolása azonban csak a fogalom bevezetésével vált lehetővé: ezeket az embereket nem megtaláltuk, hanem létrehoztuk. Ugyanígy az is elképzelhető, hogy a fogalmaink megváltoznak, és a személyiségzavart nem tekintjük többé használható kategóriának – ez esetben viszont a Kripke által felvetett újság cím a szükséges módosításokkal érvényes lenne: „Kiderült, hogy nincs is többszörösen összetett személyiség. Amit eddig annak gondoltunk, az valójában nem is az.”<sup>26</sup>

<sup>25</sup> Hacking 1995, illetve 1999. 125–162.

<sup>26</sup> Az eredeti, arannyal kapcsolatos felvetést lásd Kripke 2007. 91.

Természeti fajták esetében ugyanezt nem tudjuk megtenni: ha a tudomány jelenlegi feltevései és kategóriái tökéletesen valótlannak bizonyulnának például a bálnák természetére vonatkozóan, az sem jelentené, hogy bálnák nem léteznek – csupán azt, hogy nem olyanok, amilyennek hittük őket. A természeti világ sajátosságainak felfedezésében és fajtáinak csoportosításában ugyanúgy tévedhetünk, ahogyan a társadalmi fajták esetében is félreérthetjük a valóságot – a két tévedés azonban eltérő következményekkel jár. Előbbi esetben téves képet alkotunk az adott jelenség bizonyos tulajdonságairól, utóbbiban azonban olyasvalamiről alkotunk képet, ami *nem is létezik*. Dupré és – kisebb mértékben – Kitcher pluralista megközelítése képes számot vetni ezzel a lehetőséggel, azonban a biológiai törvények magyarázóerejét már az ő megfontolásai is csökkentik – a társadalmi jelenségek fent elemzett sajátosságai pedig még komolyabb kételyeket ébreszthetnek az azokat biológiai sajátosságokra redukálni kívánó naturalista tudománykonceptióval kapcsolatban.

Egy redukcionista vállalkozásnak ugyanis számot kellene adnia a redukció lépcsőfokairól, a társadalmi fajtákról biológiaiakra történő átmenet viszont olyasvalamit involvál, amit az utóbbi által lehetővé tett magyarázatok és előrejelzések nyelvére nem lehet leképezni. A vizsgált jelenségek fogalmi meghatározottsága az átmenet során nem pusztán módosul, hanem egyenesen eltűnik. Bármelyik megközelítést fogadjuk is el ugyanis – és bármelyikről derüljön is ki akár, hogy az az egyedüli helyes felfogás –, a biológiai fajok elhatárolására, a kategorizálendő jelenségek léte vagy nemléte egyik esetben sem tőlünk függ.<sup>27</sup> A tudományos magyarázat, az indukció és az előrejelzés valami – legalább – rövid távon változatlant feltételez ahhoz, hogy működni tudjon, a társadalmi jelenségek esetében viszont a fogalmiság meghatározó szerepe kétségesé teszi ezek sikerességét. A társadalmi fajtákra vonatkozó teoretikus általánosítások túlságosan törekénynek tűnnek: a változásokra reagálva a társadalomtudományok ugyan létrehozhatnak teljesen új kategóriákat is, ám ezek a kategóriák a tudományos bizonyosság jóval alacsonyabb fokát garantálhatják csak, mivel állandó interakcióban vannak az általuk kategorizált emberekkel, és fogalmaink megváltozása elegendő ahhoz, hogy érvénytelenné váljanak. A predikcióval kapcsolatban a helyzet még nehezebb: ez esetben ugyanis a hangsúly épp a fajtáink változatlanságára helyeződik. Érvényes és sikeres predikciók olyasvalamivel kapcsolatban tehetők, ami nincs kitéve egy esetleges hirtelen változásnak – társadalmi fajtáinkról ezt azonban nem lehet kijelenteni. Ahhoz, hogy a természettudományos módszerrel tett prognózisaink hitelesek legyenek a társadalmi jelenségekkel kapcsolatban is, nem pusztán azt kellene elérnünk, hogy az adott kategóriáink alapot adhassanak

<sup>27</sup> Természetesen létrehozhatunk új fajokat keresztezéssel, illetve tenyésztéssel, ezek azonban továbbra sem gondolhatók el a társadalmi fajok mintájára, mivel nem pusztán azáltal léteznek, hogy fogalmat alkotunk róluk.

induktív általánosításokra, hanem azt is, hogy segítségükkel előre lássuk esetleges megváltozásukat.

A társadalmi jelenségek tehát fogalmi meghatározottságuk és jelentéstelített-ségük miatt alkalmatlanok a természettudományos módszerekkel történő általánosítások és előrejelzések megbízhatóságának biztosítására, valamint a pusztán biológiai terminusokkal történő értelmezésekre – függetlenül attól, hogy milyen elméletek alapján húzzuk meg a természeti fajták határvonalait.

## IRODALOM

- Boyd, Richard 1991. Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds. *Philosophical Studies*. 61. 127–148.
- Boyd, Richard 1999. Kinds, Complexity and Multiple Realization. *Philosophical Studies*. 95. 67–98.
- Dupré, John 1981. Natural Kinds and Biological Taxa. *The Philosophical Review*. 90/1. 66–90.
- Dupré, John 1999. Are Whales Fish? In Douglas L. Medin – Scott Atran (szerk.) *Folkbiology*. Cambridge, Massachusetts, MIT Press. 461–476.
- Ereshefsky, Mark 2008. Systematics and Taxonomy. In Sahotra Sarkar – Anya Plutynski (szerk.) *A Companion to the Philosophy of Biology*. Oxford, Basil Blackwell. 99–118.
- Hacking, Ian 1986. Making up People. In Thomas Heller – Morton Sosna – David E. Wellberry (eds.) *Reconstructing Individualism*. Stanford, Stanford University Press. 222–237.
- Hacking, Ian 1995. *Rewriting the Soul*. Princeton, Princeton University Press.
- Hacking, Ian 1996. The Disunities of the Sciences. In Peter Galison – David J. Stump (szerk.) *The Disunity of Science*. Stanford, Stanford University Press. 37–74.
- Hacking, Ian 1999. *The Social Construction of What?* Cambridge, Harvard University Press.
- Hull, David L. 1965. The Effect of Essentialism on Taxonomy – Two Thousand Years of Stasis (I). *The British Journal for the Philosophy of Science*. 15/60. 314–326.
- Kitcher, Philip 1984. Species. *Philosophy of Science*. 51/2. 308–333.
- Kripke, Saul 2007. *Megnevezés és szükségszerűség*. Budapest, Akadémiai Kiadó.
- Millikan, Ruth Garrett 1999. Historical Kinds and the “Special Sciences”. *Philosophical Studies*. 95. 45–65.
- Putnam, Hilary 1975. The Meaning of “Meaning”. In *Mind, Language and Reality – Philosophical Papers*. 2. Cambridge, Cambridge University Press. 215–271.
- Reydon, Thomas A. C. 2006. Generalizations and Kinds in Natural Science: The Case of Species. *Studies in History and Philosophy of Biological and Biomedical Sciences*. 37. 230–255.
- Suppes, Patrick 1993. The Plurality of Science. In uő *Models and Methods in the Philosophy of Science: Selected Essays*. Dordrecht, Kluwer. 41–54.
- Wilkerson, T. E. 1998. Recent Work on Natural Kinds. *Philosophical Books*. 39/4. 225–233.

## Naturalizmus, szkepticizmus és racionalitás

### Plantinga evolúciós szkeptikus érve

Alvin Plantinga egyike a kereszténység (illetve a teizmus) legelkötelezettebb és legszofisztikáltabb védelmezőinek az analitikus filozófia világában. Amíg az általa kidolgozott „református episztemológia” alapvetően „defenzív” filozófiai vállalkozás (azt kívánja kimutatni, hogy a keresztény teizmus nem kell, hogy bármiféle külső megalapozásra támaszkodjék),<sup>1</sup> addig az elmúlt másfél évtizedben Plantinga „offenzívába” ment át; az általa kidolgozott, és sok vitát kiváltott „evolúciós érv a naturalizmussal szemben” (*evolutionary argument against naturalism*; EAAN) már azt kívánja kimutatni, hogy a naturalizmus<sup>2</sup> önmagában inkoherens, önmagát cáfolja – mert a naturalistának saját világnézete alapján nincs oka arra, hogy bízzék abban: kognitív képességei igaz hiteket eredményeznek, s éppen ezért minden hite kétségessé kell, hogy váljék számára – beleértve magát a naturalizmust is.

Plantinga szerint az ember és az emberi kognitív képességek evolúciós, biológiai eredete egyfajta megcáfoló megfontolás („*defeater*”) a naturalizmussal szemben.<sup>3</sup> A megcáfoló megfontolások működését Plantinga a következő példa alapján gondolja el: képzeljünk el egy keresztény hívőt, aki megtudja – például marxista vagy freudista valláseméleti munkákat olvasva –, hogy a vallásos hit valamiféle „vágykielégítő” szerepet tölt be az emberi személyiség működésében, és aki egyébként tudja azt, hogy a vágyainkon alapuló, lelki szükségleteinket kielégi-

<sup>1</sup> A „református episztemológia” ismertetését és kritikáját magyar nyelven lásd Szalai 2005.

<sup>2</sup> A „naturalizmus” mint filozófiai terminus használata természetesen magában hord bizonyos többértelműséget. De Plantinga használatában a kifejezés alapvetően arra az álláspontra vonatkozik, hogy Isten vagy más természetfeletti, immateriális lények nem léteznek, s a naturalizmus mibenlétének pontos definiálásával kapcsolatos viták nem érintik a jelen tanulmányban vizsgált kérdéseket.

<sup>3</sup> Plantinga az érvet többször is kifejtette, először 1993-ban, *Warrant and Proper Function* című munkájában. A jelen vázlat Plantinga egy nem publikált, egyedül az interneten hozzáférhető, ámde a szakirodalomban széles körben ismert és vitatott 1994-es cikkén alapszik. Lásd Plantinga 1994.

tő gondolkodás általában nem megbízható kalauzunk az objektív igazsághoz. Ez a teista megtudott valamit hitei, illetve kognitív képességei *eredetéről*, ami megcáfoló megfontolásként, „defeater”-ként fog működni a továbbiakban, ami azt jelenti, hogy nem racionális többé hinnie a vallásban.<sup>4</sup> Hasonló megcáfoló megfontolás lesz Plantinga szerint az evolúcióelmélet a naturalista számára. De vajon miért?

Plantinga abból indul ki, hogy a természetes kiválasztódás jutalmazza az adaptív viselkedést, s megbünteti a nem alkalmazkodót, de egyáltalán nem törődik azzal, hogy igaz-e, amit hiszünk. Idézi Patricia Churchlandot, a kortárs naturalista elmefilozófia kiemelkedő gondolkodóját, aki szerint az evolúció során kialakuló idegrendszeri struktúrák célja elsősorban az, hogy az organizmusokat sikeressé tegyék az alkalmazkodás során legfontosabb négy alapvető feladat: a táplálkozás, a természetes ellenségek előli menekülés, a természetes ellenségekkel (verseny társakkal) való megküzdés és a szaporodás elvégzésében (Churchland 1987). Kognitív képességeink (a naturalizmus szerint) az idegrendszer funkciói, tehát nyilván alkalmasnak kell lenniük arra, hogy elősegítsék e célok megvalósítását. Ámde az, hogy eme alkalmazkodási célok megvalósításától függetlenül mennyire megbízhatóan ismerik meg a valóságot, az evolúció szempontjából irreleváns. Milyen módon befolyásolja ez a tény kognitív képességeink megbízhatóságát? Legyen eme képességeink megbízhatóságának hipotézise R; legyen a materializmus/naturalizmus igaz voltának hipotézise N, és legyen az evolúció hipotézise E, végül jelöljük a valószínűséget P-vel. Mekkora lesz P (R/N&E)?

Plantinga szerint ezt a kérdést először egy olyan gondolatkísérlet formájában kell végiggondolnunk, amelynek során nem rólunk, emberekről, hanem egy másik bolygón élő feltételezett fajról van szó, akikről tudjuk, hogy a mieinkhez hasonló értelmi funkciókkal bírnak, és hogy az emberi evolúcióhoz hasonló természeti folyamatok nyomán fejlődtek ki, és tudnak alkalmazkodni saját bolygójuk viszonyaihoz. Viselkedésük legnagyobb részét nyilván adaptív (különben nem tudtak volna bolygójukon fennmaradni), de mi a helyzet a hiteikkel, valószínű-e hogy ezek legnagyobb részt igazak? Plantinga szerint, ha az idegen bolygón élő, feltételezett értelmes faj esetében igazoljuk, hogy P (R/N&E) ismeretlen vagy alacsony, akkor ez megáll ránk is. Ennek eldöntéséhez meg kell vizsgálnunk azt a kérdést: hogyan függhetnek össze ezeknek a feltételezett, racionális lényeknek a hitei a viselkedésükkel?

Plantinga szerint négy, egymást kölcsönösen kizáró és együttesen a lehetőségeket kimerítő elméleti alternatívánk lesz ezzel kapcsolatban:

(1) Az egyik lehetőség az *epifenomenalizmus*: az értelmes lények viselkedését egyáltalán nem a hiteik okozzák. E hipotézis szerint mozgásukat és viselkedé-

<sup>4</sup> Természetesen nem minden „defeater” vonatkozik hiteink, illetve kognitív képességeink *eredetére*; „defeater” lehet például egy olyan érvelés is, amely szerint egy meghatározott hitem inkonzisztens egy másik, az elsőnél valószínűbbnek tartott hittel.

süket más tényezők – például idegrendszeri impulzusok – irányítják, amelyeket más szervezetükben fennálló tényállások, például az érzékszerveket ért ingerek okoznak, de a *hitnek* nem lesz helye ebben a viselkedéshez vezető kauzális láncban. A viselkedés és a hit (valamint a többi mentális jelenségek, mint amilyen az érzés, érzékelés és vágy) közötti viszonynak ez a felfogása ma meglehetősen népszerű, különösen a biológusok körében. Végül is a biológusok úgy hiszik, hogy a viselkedést, bármilyen komplex is, teljesen a biokémia szabja meg, hogy az ezt kísérő érzések – a félelem, a fájdalom, a csodálkozás, a szeretet – egyszerűen csak „olyanok, mint az árnyékok a fizikai tárgyakhoz képest, tehát önmagukban lényegében nem érintik az organizmus viselkedését...”<sup>5</sup> A biológiai „ortodoxia” szerint ugyanez vonatkozik a hitekre is – feltételezve, hogy a hitek önmagukban nem *pusztán* biokémiai jelenségek. Hogyha ez a felfogás a gondolatkísérletünk által feltételezett lényeket illetően helytálló, akkor hitek *láthatatlanok* lesznek az evolúció számára, és ekkor az a tény, hogy hitek-kialakító mechanizmusaik evolúciójuk története során alakultak ki, kevéssé vagy egyáltalán nem teszi valószínűvé, hogy hitek legnagyobbbrészt igazak. N&E és az első lehetőség alapján tehát  $P(R)$  meglehetősen alacsony lesz.

(2) Egy második lehetőség a *szemantikai* epifenomenalizmus: lehetséges az is, hogy ezeknek a lényeknek a hitei ugyan ok-okozati viszonyban állnak a viselkedésükkel, ámde nem *tartalmuk* révén. Ez az elképzelés azt jelentené, hogy hitek rendelkeznek ugyan kauzális hatékonysággal, de neurofiziológiai beágyazottságuk, tulajdonságaik, nem pedig *szemantikai tulajdonságaik* révén. A naturalista vagy bármilyen értelemben materialista gondolkodásmód szerint egy hit feltehetően az idegrendszeri aktivitás valamilyen hosszú távú „mintája”. Hosszú távú neuronális történés. Ez az esemény legalább kétfajta tulajdonsággal rendelkezne. Egyfelől lennének elektrokémiai tulajdonságai: a hit létrejöttében érintett neuronok száma, a közöttük fennálló kapcsolatok, külsülési küszöbük, a külsülések üteme és erőssége, az, ahogyan ezek bizonyos idő alatt, s más neuronális működésekkel összefüggésben változnak, és így tovább. Másfelől azonban ha a hit valóban *hit*, akkor (valamely  $p$  hitet illetően) az a hit lesz, *hogy p*. Például az a hit lesz, hogy egyszer egy söröző állt a Magyar Állami Operaház helyén. Ez a kijelentés lesz a kérdéses hit *tartalma*. Így egy hitnek neurofiziológiai tulajdonságain túlmenően *szemantikai* tulajdonságai is lesznek – például az, hogy ez a hit *éppen az a hit, hogy* „egyszer egy söröző állt a Magyar Állami Operaház helyén”. (Más szemantikai tulajdonságok: *igaznak vagy hamisnak lenni, implikálni, hogy volt legalább egy söröző Budapesten, konzisztensnek lenni azzal a kijelentéssel, hogy minden ember halandó* és így tovább.) A szemantikai epifenomenalizmus mármint az a felfogás, hogy a hit ugyan kauzálisan hatékony a viselkedést illetően, azonban *a neurofiziológiai*, nem pedig szemantikai tulajdonságai révén. Ha az első felfogás

<sup>5</sup> J. M. Smith (*Time*, December/1992), idézi Plantinga 1994. 7.

népszerű a biológusok, akkor ez utóbbi viszont széles körben népszerű a kortárs elmefilozófusok körében.

E felfogás alapján P(R/N&E) ezekre a lényekre specifikálva éppúgy alacsony lesz, mint az első nézet alapján. Ennek az indoka az, hogy az igazság vagy hamisság természetesen a hit szemantikai, nem pedig *neurofiziológiai*<sup>6</sup> tulajdonságai közé tartozik. De ha az előbbieket nem vesznek részt a hitekhez vezető oksági láncban, akkor a hitek – vagyis inkább azok szemantikai tulajdonságai, beleértve az igazságot és a hamisságot – megint csak láthatatlanok lesznek a természetes kiválasztódás számára. De akkor valószínűtlen lesz, hogy e lények hitei legnagyobb részben igazak, s ezért valószínűtlen az is, hogy kognitív képességeik megbízhatóak. R valószínűsége N&E plusz eme lehetőség alapján ezért relatíve alacsony lesz.

(3) Elképzelhető az, hogy a hitek kauzálisan hatékonyak – „szemantikailag” éppúgy, mint „neurofiziológiailag” –, ámde *negatívan* érintik az alkalmazkodást: az alkalmazkodóképesség szempontjából ezek a lények jobban boldogulnának a hitek *nélkül*. Nyilvánvalóan R valószínűsége N&E-nek és ennek a lehetőségnek az alapján, akárcsak az előző két esetben, szintén alacsony lenne.

(4) Végül lehetséges az is, hogy elképzelt lényeink hitei kauzálisan összekapcsolódnak viselkedésükkel, és egyszersmind adaptívak is. (Vélhetően ez a viselkedés és a hit közötti összefüggés „common sense” felfogása.)

Mi a valószínűsége annak (eme hipotézis plusz N&E alapján), hogy a lények kognitív képességei megbízhatóak, illetve az e képességek által létrehozott hitek igazak lesznek? Plantinga szerint ez a valószínűség távolról sem olyan magas, mint amilyenek első megközelítésben gondolnánk. Ennek az oka az, hogy ha a viselkedést a *hitek* idézik elő, akkor egyszersmind előidézi a *vágy* is (valamint az összes többi itt ignorált pszichés tényező, például: gyanakvás, kétség, helyeslés és helytelenítés, félelem, stb). Bármely adaptív viselkedés vagy cselekvés esetén sokféle hit-vágy kombináció lesz, amely ehhez a cselekvéshez vezethet; s eme hit-vágy kombinációk közül nagyon sok olyan lesz, amelyben a bennük szereplő hit hamis.

Tételezzük fel például, hogy Paul egy történelem előtti korban élt ősember, akihez egy éhes tigris közeledik. A legmegfelelőbb viselkedés ebben a helyzetben valószínűleg az, ha Paul elszalad a tigris elől: ámde Plantinga kimutatja, hogy ezt a viselkedést egy sor *különböző* hit-vágy kombináció okozhatja. Paul például nagyon is vágyik arra, hogy vadállatok megegyék, de amikor meglát egy tigrist, akkor mindig elszalad, mert valamiért valószínűtlennek tartja, hogy a tigris meg fogja őt enni. Ez az érzelmi reflex mindig az életben maradásához szükséges módon fogja irányítani Paul testét, ámde ennek semmi köze sincs Paul

<sup>6</sup> Plantinga a hiteknek ezekre a nem-szemantikai tulajdonságaira a „szintaktikai” megjelölést használja, azonban ezt a szóhasználatot tanulmányomban mellőzöm, mivel csak értelméleti zavarokhoz vezetne.



igaz hiteihez. Vagy Paul azt gondolja, hogy a tigris egy barátságos, nagy macska és meg akarja szelídíteni, egyszersmind úgy hiszi, hogy a legjobb módszer erre az, ha játékosan fogócskázni kezd a macskával, és elszalad, amikor látja. Vagy tegyük fel, hogy Paul törzsében futóversenyt rendeznek, Paul szeretne nyerni a versenyben, és azt hiszi, hogy a tigris megjelenése adja meg a startjelet... Nyilvánvalóan nagyon sokféle hit-vágy kombinációt lehet elképzelni, amely egy adott viselkedésformát (jelen esetben azt, hogy Paul elszalad a tigris elől) eredményez.

Itt felmerül egy probléma a bemutatott érveléssel kapcsolatban. Könnyű belátni, hogy Paul *egyik* cselekvését sokféle különböző vágy-hit kombináció eredményezheti; de nehezebb megérteni/elképzelni azt, hogy miként lehetne *összes* hiteinek legnagyobb része hamis oly módon, hogy közben mégis adaptív következményekkel jár. Plantinga szerint a legegyszerűbb módon úgy láthatjuk be ennek a lehetőségét, ha elgondolunk olyan, a gondolkodást szisztematikusan félrevezető „utakat”, amelyekeken haladva rendszerint hamis, de mégis adaptív viselkedéseket alakíthat ki valaki. Lehet például Paul valamiféle primitív animizmus híve, és gondolhatja azt, hogy minden dolog élő. Talán azt gondolja, az összes növények és állatok boszorkányok, s mindig ilyen kifejezésekkel utal rájuk. Például: „ez a boszorkány jó táplálék”, „ez a boszorkány közömbös az ember iránt”, „ez a boszorkány veszélyes”. Paul hitei ez esetben kivétel nélkül hamisak lesznek, de ebből nem következik, hogy nem megfelelő (nem adaptív) módokon fog viselkedni a környezetében található állatok és növények irányában.

Amit eddig láttunk, az az, hogy négy egymást kölcsönösen kizáró és a lehetőségek teljes skáláját kimerítő lehetőségünk van az idegen bolygó feltételezett lakosait illetően: az egyszerű epifenomenalizmus, a szemantikai epifenomenalizmus, az a lehetőség, hogy hiteik tartalmi kauzálisan hatnak ugyan viselkedésükre, de negatívan érintik alkalmazkodási esélyeiket, s az a lehetőség, hogy hiteik kauzálisan hatékonyak, s adaptívak.  $P(R/N\&E)$ -t úgy számolhatjuk ki, hogy felbecsüljük  $P(R/N\&E)$ -t e hipotézisek mindegyikét illetően, majd beszorozzuk ezeket a valószínűségeket mindegyik hipotézis  $N\&E$  melletti valószínűségével, s az így nyert értékeket összeadjuk. (Természetesen nevetséges lenne valóban pontos számokat követelnünk, hiszen e valószínűségeket illetően csak homályos becsléseink lehetnek.)

Tehát (a négy vázolt filozófiai lehetőséget – hipotézist –  $H_n$ -vel jelölve):

$$P(R/N\&E) = (P(R/N\&E\&H1) \times P(H1/N\&E)) + (P(R/N\&E\&H2) \times P(H2/N\&E)) + (P(R/N\&E\&H3) \times P(H3/N\&E)) + (P(R/N\&E\&H4) \times P(H4/N\&E))$$

A fenti megfontolások arra mutatnak, hogy  $P(R/N\&E)$  az első három hipotézis esetén elég csekély lesz, a negyedik esetén pedig elég magas ugyan, de nem nagyon magas, nem sokkal nagyobb 0,5-nél. (Mert – miként a fenti példák mutatják – a hitek akkor is eredményezhetnek adaptív viselkedést, ha nem igazak.) Ugyanakkor látnunk kell, hogy a naturalizmus mellett  $H_1$  és  $H_2$  meglehetősen valószínűek – mert naturalista alapon nem nagyon tudjuk elgondolni, hogy a hitek tartalma miként lehet hatással az organizmusok viselkedésére. Ha most ezeket a becsléseket behelyettesítjük a képletbe, akkor mindenképpen azt fogjuk kapni, hogy  $P(R/N\&E)$  kisebb lesz 0,5-nél – a naturalizmust és az evolúciót feltételezve nem túl valószínű (kevésbé valószínű, mint az ellenkezője), hogy az idegen bolygó értelmes lények kognitív képességei megbízhatóak.

De természetesen valaki lehet szkeptikus is e valószínűségek mindegyikével kapcsolatban, s ezen az alapon állíthatja azt, hogy  $P(R/N\&E)$  *meghatározhatatlan*.

Plantinga érvének végső konklúziója tehát az, hogy  $P(R/N\&E)$  alacsony vagy meghatározhatatlan. De ha ez így van, akkor szerinte ez nemcsak az idegen bolygó lakóira, hanem *miránk magunkra* is igaz: saját kognitív képességeinkről is azt kell gondolnunk, hogy ha a naturalizmus és a fejlődésemélet igaz, akkor megbízhatóságuk nem túl valószínű. Ezért az is valószínű, hogy hiteink nagy része nem igaz.<sup>7</sup> Tehát nincs különösebb okunk megbízni egyetlen hitünkben sem, beleértve magát a naturalizmust is. N&E ezért a naturalisták számára (ha végiggondolják azokat az implikációit, amelyeket Plantinga a fentiekben vázolt) egy megcáfoló megfontolás („*defeater*”), egy olyan megfontolás, amely a naturalizmusba vetett hitüket irracionálissá teszi.

Természetesen a megcáfoló megfontolásokat magukat is meg lehet cáfolni. Például ha én azt hiszem, hogy Béla tud úszni, akkor egy megcáfoló megfontolás hitemre nézve, hogy Béla egy olyan településen él, ahol a lakosok 90%-a nem tud úszni. Ámde ezután tudomást szerzek arról, hogy Béla a szóban forgó településen az uszodában dolgozik, márpedig a magyar uszodák alkalmazottainak 99%-a tud úszni. Második hitem ez esetben megcáfolja az első hitemmel szembeni megcáfoló megfontolásaimat. Csakhogy Plantinga szerint a naturalista azért nem találhat ilyen megcáfoló megfontolást megcáfoló megfontolást („*defeater-defeater*”) a jelen esetben, mert ez csak valamilyen további *hite* lehetne.

<sup>7</sup> Meggyőződésem szerint Plantinga figyelmen kívül hagy a gondolkísérletben egy nagyon fontos kérdést: vajon ezek az értelmes lények az *általuk értelmi képességeik tudatos alkalmazása révén létrehozott technológiák* jóvoltából alkalmazkodnak-e saját bolygójuk viszonyaihoz (ahogyan mi tesszük a Földön), vagy sem? Ez a kérdés nagymértékben meghatározza a gondolkísérlet eredményeit. Annak a valószínűsége ugyanis, hogy értelmes lények esetleg olyan alkalmazkodó viselkedésformák segítségével maradnak fenn egy adott környezetben, amelyeknek semmi közük hiteink igazságtartalmához, hanem mintegy „ösztonösen” alakultak ki és működnek, lehet viszonylag nagy. Annak a valószínűsége viszont, hogy értelmes lények, akik legnagyobb részt téves hiteket vallanak környezetükről, ezeknek a téves hiteknek az alapján ezt a környezetet sikeresen módosító és saját céljaikra felhasználó technológiákat tudnak feltalálni és működtetni, elenyészőnek tűnik.

Viszont az R-re vonatkozó kételyünk *minden* hitünkre vonatkozóan egy megcáfoló megfontolás; ha nem racionális elfogadnunk R-t, akkor nem racionális elfogadnunk *egyetlen* hitünket sem.

Plantinga szerint tehát aki *elfogadja* N&E-t, az egyszersmind rendelkezik egy *megcáfoló megfontolással* N&E-t illetően, egy indokkal arra, hogy kétségbe vonja vagy elvesse. N&E tehát önmagát cáfolja, s ezért elfogadása irracionális. Plantinga megfogalmazása szerint az evolucionista naturalista<sup>8</sup> olyan helyzetben van, mint az az ember, aki tudja, hogy elfogyasztott egy olyan X gyógyszert, amely kognitív képességeit (mondjuk) 90%-ban megbízhatatlanná teszi. Az ilyen embernek nem lenne oka arra, hogy *bármely* hitében megbízzon.

A teista (mohamedán, zsidó vagy keresztény) hívő viszont hiszi, hogy Isten teremtett bennünket, saját képmására, amihez hozzátartozik (többek között) az is, hogy sok mindent tudhatunk és tudunk is a körülöttünk lévő világról. Ezért jó indokai vannak arra a meggyőződésére, hogy Isten az evolúciót úgy irányította, hogy az értelmes, a valóságot megismerni tudó lényeket hozzon létre. A teista világképe így egyrészt koherens lesz a fejlődésemeléttel, másrészt pedig nem cáfolja önmagát, mint a naturalizmus. S ez természetesen döntő indokunk lehet arra, hogy preferáljuk a naturalizmussal szemben.<sup>9</sup>

Plantinga érvelése, mint láttuk, alapvetően három lépésből áll. 1. Az emberi kognitív képességek kialakulásának naturalista, evolúciós „története” valószínűvé teszi, hogy e képességeink megbízhatatlanok. 2. Ezért minden okunk megvan rá, hogy ne bízunk meg egyetlen hitünkben sem; tehát magában a naturalizmusban és a fejlődéseméletben sem. 3. Ebből a szkeptikus zsákutcából csak a teizmus elfogadása „szabadíthat ki” bennünket. Mindhárom lépésnek számos kritikája ismeretes.<sup>10</sup> Megkérdőjelezhető az a leírás, amelyet Plantinga a hitek szemantikai és neurofiziológiai tulajdonságairól és a köztük lévő viszonyról ad. De ezt elfogadva is vitatható marad, hogy a biológiai organizmusokban (köztük az emberekben) az adaptációs előnyök és az igaz hitekkel/megbízható kognitív képességekkel való rendelkezés csakugyan annyira függetlenek lennének egymástól, ahogyan Plantinga állítja. Attól ugyanis, hogy az igaz hitek, illetve megbízható kognitív képességek és az adaptáció közötti kapcsolatot nem tudjuk *megmagyarázni*, filozófiailag elemezni, még teljesen elképzelhető, hogy *ez az összefüggés igazolhatóan fennáll*.

<sup>8</sup> Tisztán elvben a naturalista Plantinga támadásával szemben úgy is védekezhetne, hogy feladja az evolúcióba vetett hitét. Gyakorlatilag azonban, mint Plantinga rámutat, az emberi lények kialakulásának egyedüli valamennyire is plauzibilis naturalisztikus magyarázata a fejlődésemélet.

<sup>9</sup> Plantinga szerint egyáltalán nincs feszültség a teizmus és a fejlődésemélet között, hanem csak az evolúció és a Biblia fundamentalista értelmezése között. Máshol megpróbáltam amellette érvelni, hogy ez nem így van, de a jelen tanulmányban nem térek ki erre. Lásd Szalai 2009.

<sup>10</sup> Megkíséreltem az összes kritikát vázlatosan ismertetni és kiértékelni a *Magyar Tudomány*-ban megjelent tanulmányomban. Lásd Szalai 2009.

Ámde ha mindezt meg is engedjük Plantingának, ha az emberiség evolúciójának története (önmagában) csakugyan egy szkeptikus végkicsengésű „forgatókönyvet” adna is a kezünkbe saját kognitív képességeink eredetéről és működéséről, akkor is kérdés: elégséges-e ez ahhoz, hogy megrendítse a megismerésbe vetett eredendő bizalmunkat, illetve kihúzza az alapot ismereteink egész rendszere alól? A jelen tanulmányban csak ez utóbbi kérdéssel, tehát Plantinga érvének *episztemológiai* aspektusaival kívánok foglalkozni – megengedve Plantingának, hogy az érvelés (1) lépése sikeres. (Az érveléssel szemben elképzelhető egy negyedik típusú, a szakirodalomban *tu quoque*-nek nevezett ellenvetés is, amely szerint akár alátámasztja az EAAN a megismeréssel szembeni szkepszist, akár nem, a *teista* semmivel sem lesz könnyebb helyzetben a szkeptikussal szemben, mint a naturalista. Ennek az ellenvetésnek is többféle megfogalmazása lehetséges – tanulmányom végén röviden visszatérek még a kérdésre.)

Az episztemológiai jellegű kritikák az EAAN-nal szemben legalább három-féleképpen érvelnek. Az első típusú érvelés szerint az EAAN egész egyszerűen inkoherens, hiszen az emberi kognitív képességeket *alkalmazva*, azok megbízhatóságát feltételezve érvel a teljes szkepticizmus mellett. A második típusú (*common sense*, illetve externalista) érvelés szerint az EAAN hibás módon következtet: abból, hogy az evolúció *nem támasztja alá* ismereteink megbízhatóságát, nem következtethetünk arra, hogy alátámasztaná ismereteink *nem-megbízhatóságát* (Fodor 2002. 30–42). Ez az alapvető bizalmunk ugyanis magától értetődő számunkra – amíg filozófiai kételyek meg nem rendítik. *Egyes* ismereteinkben indokoltan kételkedhetünk azért, mert nincsenek összhangban ismereteink rendszerének egészével, de a rendszer *egészére* vonatkozó kételyek indokolatlanok és abszurdak. A harmadik típusú ellenvetés szerint *elvéileg* lehetnek jó átfogó szkeptikus érvek az emberi megismerés rendszerének egészével kapcsolatban, de Plantinga érve nem ilyen. Ezt az utóbbi, a szkeptikust elvileg vitapartnernek tekintő választ fogom „kartezianus” válasznak nevezni.

Az alábbiakban megkíséreltem kiértékelni e három lehetséges kritika mind-egyikét.

## I. INKOHERENS-E AZ EAAN?

Több filozófus, mindenekelőtt Thomas H. Reed próbálkozott Plantinga érvelésének olyan kritikájával, amely szerint az EAAN egyszerűen önmagát cáfolja (Reed 1997). Ugyanis az a naturalista, aki elfogadja az EAAN nyomán azt, hogy  $P(R/N\&E)$  alacsony, kénytelen összes kognitív képességeinek csekély megbízhatóságot tulajdonítani, kénytelen azt gondolni, hogy legtöbb hite nem igaz; ámde ha ez így van, akkor ez vonatkozik azokra a kognitív képességeire is, amelyek az EAAN érvényes voltának belátásához elvezették. Következésképpen vagy nem megbízhatóak kognitív képességeink, és akkor jó indokkal kételke-

dünk magában az EAAN-ban is, vagy pedig megbízhatóaknak bizonyultak ebben az esetben – de akkor jó eséllyel feltételezhetjük, hogy általában is azok – és akkor az EAAN megint csak nem lesz érvényes. Van-e Plantinga számára kiút ebből a dilemmából?

Troy Nunley szerint – aki egy teljes disszertációt írt az EAAN-ról – van (Nunley 2005). Nunley szerint Reed abból a – védhetetlen – álláspontból indul ki, hogy amikor egy A személy egy B személy ellen érvel, akkor meg kell engednie B-nek mindazokat a feltevéseket, amelyeket ő maga is vall vagy feltételez. Ha ez így volna, akkor a külvilágra vonatkozó szkeptikus érvek vagy a „más tudatokra” (other minds) vonatkozó szkepszis hangoztatóival szemben is nagyon könnyű helyzetben volnánk. Egyszerűen megkérdezhetnénk tőlük: ők maguk léteznek-e, vagy nekik maguknak van-e tudatuk? Mivel ezekre a kérdésekre nyilván „igen”-nel fognak válaszolni, máris posztulálhatunk egy saját elménktől különböző objektumot, vagy egy elmét – és ezzel a külvilágra vagy a más tudatokra vonatkozó szkepszis el lenne intézve. Ámde nyilvánvalóan ez nem megfelelő érvelés a szkeptikusokkal szemben, akkor viszont Reed érvelése sem lehet megfelelő stratégia Plantinga ellenében.

Hogyha Reednek (vagy más filozófusnak) valóban olyan módon kellene érvelnie Plantinga érvének inkoherenciája mellett, ahogy Nunley véli, akkor ügyük valóban eleve elveszett volna. Valamely érvelés ellen nem lehet felhozni ellenérvként, hogy ellentmond valaminek, amit *az, aki az érvet kifejti*, állít. Egy érvelés csak akkor inkoherens, ha maga az érvelés feltételez, vagy implicál valamit, ami konklúziójának ellentmond. De feltételez-e az EAAN valamilyen olyan premisszát, amely ellentmond konklúziójának?

Önmagában az, hogy az EAAN az emberi kognitív képességek megbízhatóságának hipotéziséből indul ki, majd ahhoz a konklúzióhoz jut el, hogy ezek a képességek végső soron nem megbízhatóak, egyáltalán nem jelenti azt, hogy az EAAN ellentmondana önmagának. Hiszen a filozófiában teljesen szokványos abból kiindulva, hogy  $p$  amellett érvelni, hogy  $\sim p$ . Úgyszintén nem inkoherens az EAAN pusztán azért, mert annak, aki nyomon követi lépéseit, az *érvelés során* fel kell tételeznie kognitív képességei (legalábbis azok, amelyek az érvelés megértéséhez szükségesek) megbízhatóságát. Ugyanis Plantinga expliciten állítja, hogy a naturalisták az érvelés *követése* során igenis fenntarthatják kognitív képességeik megbízhatóságára vonatkozó meggyőződésüket, egyedül az érvelés *végére érve*, az érv konklúziójához eljutva kell hogy feladják azt. Az más kérdés, hogy ha egyszer feladják, akkor utána újra elkezdenek kételkedni az EAAN-ban – de ebben az esetben újakezdhetik a „játszmát”, újból feltételezve először N&E-t, és az emberi kognitív képességek megbízhatóságát, majd levonva ezekből az előfeltevésekből a megfelelő szkeptikus konklúziókat. Természetesen Plantinga is tisztában van vele, hogy a normális emberek ez esetben egy idő után nyilván „feladnák” a kérdéssel való foglalkozást, s ugyanúgy visszatérnének a normál kognitív működéseikbe vetett bizalomhoz, ahogyan Hume is feltételez-

te, hogy az emberi tudat természettől fogva képtelen elfogadni az ő szkeptikus konklúzióit – de ez nem jelenti azt, hogy ez a lépésük filozófiailag, az igazság szempontjából is indokolt lesz.

Az EAAN meggyőződése szerint *mégis* inkoherens, a következő okból. Plantinga *nem* egyszerűen azt állítja, hogy P(R/N&E) meghatározhatatlan vagy alacsony, hanem azt is, hogy ez egy *megcáfoló megfontolást* (*defeater*) alkot a naturalisták számára, irracionálissá teszi a N-be vetett hitüket. Ámde a megcáfoló megfontolás fogalma, az a mód, ahogy ezt Plantinga definiálja, feltételezi, hogy aki egy ilyen megcáfoló megfontolással rendelkezik, vagy akár csak rendelkezhet, az rendelkezik bizonyos megbízható kognitív képességekkel, mindenekelőtt egy bizonyos mértékű rálátással saját hiteire, azok egymással való összefüggésére, továbbá rendelkezik bizonyos (legalábbis elemi) logikai készségekkel. Plantinga nem válaszolhatja azt, hogy „eleve tudja”: a naturalistákat úgysem fogja meggyőzni az ő érve – mert még ebben az esetben is feltételezi (magának a „megcáfoló megfontolásnak” a fogalmánál fogva), hogy legalábbis *elvéleg* a naturalisták beláthatják érvelése helyességét – ez viszont azt jelenti, hogy rendelkeznek bizonyos megbízható kognitív képességekkel.

Az a tény azonban, hogy ilyen módon az EAAN generálta szkepszis végül is nem tudja kikerülni az önreferenciát, nem indokolja, hogy ne vizsgáljuk tovább az érvet. A naturalizmus szempontjából filozófiailag rendkívül aggályos lenne az, ha *csak* azokat a kognitív képességeinket tekintenénk megbízhatóknak, amelyek saját ismereteink rendszerére és a köztük fennálló logikai relációkra vonatkoznak. A legtöbb naturalista szeretné azt gondolni, van némi alapunk arra, hogy a *külvilágra* vonatkozó hiteinkben és az ezeket a hiteket létrehozó kognitív képességeinkben is megbízhatunk.

## II. „COMMON SENSE” ÉS EXTERNALISTA VÁLASZOK

Általában (a tudományos vagy a hétköznapi megismerés módszereit alkalmazva) kognitív képességeink megbízhatóságában nem szoktunk kételkedni – addig, amíg a filozófiai szkepszis gondolkodóba nem ejt bennünket erről. Hogyha ez alapvető, természetes meggyőződésünk, akkor pusztán azért, mert az emberiség evolúciójára vonatkozó ismereteink nem támasztják meggyőzően alá, miért kéne feladnunk?

Plantingának azok a bírálatai, amelyek ebből indulnak ki, két csoportra oszthatóak. Az egyik csoport hívei, a „reidi” episztemológiai megközelítésre támaszkodó,<sup>11</sup> *common-sense* kritikusok abból indulnak ki, hogy a megismerés rendszerének *egészével* szemben nem lehetségesek olyan megcáfoló megfontolások, amelyek

<sup>11</sup> Thomas Reid személy szerint *teista* volt, ezért talán különösnek tűnik az ő álláspontja alapján érvelni a naturalizmus mellett. Reid azonban expliciten állította, hogy az alapvető

lehetőségek a rendszer *egyes elemeivel* szemben. E szerint az álláspont szerint számos ismeretünket (köztük „*common sense*” hiteink jelentős részét) nem más hitekből következtetjük ki, hanem valamilyen nem-propozicionális evidenciára/tapasztalatra alapozva hisszük. Éppen ezért, megengedve Plantingának, hogy P(R/E&N) az általa elképzelt idegen bolygó értelmes lakóit illetően meghatározhatatlan, vagy csekély, ebből nem kell levonnunk azt a következtetést, hogy P(R/E&N) a *mi* esetünkben is csekély, mert a saját alapvető meggyőződéseink (kognitív képességeink megbízhatóságába vetett hitünk) – ellentétben az idegen bolygó lakosaival kapcsolatos hipotéziseinkkel – nem más hiteken, illetve kijelentéseken, hanem kijelentések formájában sokszor meg sem fogalmazott tapasztalatokon (*non-propositional evidence*) nyugszanak (Bergmann 2002. 61).

A másik csoportba az externalista válasz hívei tartoznak. Eszerint az EA-AN-generálta szkepszist meg tudjuk úgy cáfolni, hogy az emberi megismerés leírása során bizonyos hiteinkre, ismereteinkre támaszkodunk. Elvégre csak az internalista ismeretelméletek szerint kell tudnunk (vagy igazoltan hinnünk), hogy tudunk (vagy igazoltan hiszünk) valamit, az externalisták szerint elég, ha a tudás (vagy az igazolt hit) kritériumai, feltételei egyszerűen *fennállnak* – akár tudunk róluk, akár nem. Ma már a legtöbb filozófus elfogadja, hogy az episztemológiai szkeptikust nem tudjuk megcáfolni. Az externalista episztemológusok szerint azonban ez nem jár semmiféle veszéllyel a tudásunkra (illetve az ezzel kapcsolatos bizonyosságunkra) nézve. Le tudjuk írni valamennyi hitünket, illetve kognitív mechanizmusunkat – és igazolni tudjuk megbízhatóságukat – úgy, hogy közben *más* hiteinket, illetve kognitív mechanizmusainkat megbízhatónak feltételezzük. Amivel nem rendelkezhetünk, az egy *egyszerre az összes* hiteinkről, illetve kognitív képességeinkről adott igazoló-megalapozó leírás – de erre törekedni nem is volna racionális (Sosa 1999. 96). Ez volna tehát Plantinga érvelésére az externalista episztemológia válasza. Semmiképpen sem *elve* kilátástalan ez a válaszkíséret, hiszen Plantinga szerint az EAAN *nem* feltételez semmilyen specifikus episztemológiai álláspontot (Plantinga 2002. 227).

Az externalista válasz abban tér el a *common-sense* választól, hogy bár elismeri: a kognitív képességeit nem-reflektált módon alkalmazó személynek természetesen igaza van, amikor mindenfajta „külső” igazolás, tudatos episztemológiai reflexió nélkül egyszerűen csak hisz a maga meggyőződéseiben – ámde azt állítja: egy, az emberek kognitív képességei és a külvilág viszonyáról szóló egységes koncepció egy magasabb szintű, jobban megalapozott tudáshoz juttathat el bennünket.

Van Cleve például Descartes példájára hivatkozik Plantingával szemben. Igaz, hogy Descartes az elmében lévő a priori fogalmak, az *idea clara et distincta* biztos, igaz ismeretet nyújtó voltában Isten igazságszeretete (*veracitas Dei*) miatt

kognitív képességeinkbe vetett természetes bizalmunk *nem* az istenhiten alapszik. Lásd Reid 1969. 294–295.

volt bizonyos, de ezt magát is az *idea clara et distincta* segítségével bizonyította. Descartes érvelése mégsem rossz értelemben vett körforgás, mert Descartes mielőtt az elméjében lévő *idea clara et distincta*-kat megvizsgálta volna, nem tudta (de még csak nem is feltételezte) hogy azok természete az igazság biztos kritériuma, megvizsgálásuk az igazsághoz vezető biztos módszer lenne.<sup>12</sup> Descartes egyszerűen csak megtudott bizonyos dolgokat az *idea clara et distincta* vizsgálatából, s ezeket a dolgokat akkor is tudná, ha nem jutott volna el oda, hogy az *idea clara et distincta* mint ismeretforrás megbízható voltáról Isten kezeskedik. Descartes expliciten elismeri, hogy egy ateista matematikus is lehet bizonyos abban, hogy  $2 \times 2 = 4$ , vagy a háromszög szögeinek összege 180 fok, ámde azt állítja, hogy az olyan matematikus, aki nyomon követve az ő érvelését eljutna Isten létezésének, a *veracitas* Deinek és így az *idea clara et distincta* mint igazságkritérium megbízhatóságának a belátásáig, egy megbízhatóbb és mélyebb tudás (*scientia*) birtokába jutna – amelyet ő szembeállít a pusztá *cognitio*-val (megismerés).

Természetesen Descartes a maga ismeretelméletéhez olyan metafizikai előfeltevéseket használt, amelyek a naturalizmussal összeegyeztethetetlenek, de *módszere* ettől még követhető marad a naturalista számára. A naturalista filozófus először – egyszerűen bizonyos kognitív képességeinket *használva*, a megbízhatóságukra vonatkozó bármiféle előzetes meggyőződés *nélkül* – kimutatja, hogy bizonyos ismereteink igazak, majd felépít ezekre az ismeretekre egy olyan teljes képet az emberi megismerés rendszeréről, amely kognitív képességeinket tudatos episztemológiai perspektívába helyezi – s így kimutatja róluk, hogy általában igenis megbízhatóak (Sosa 2002. 93–94). Ilyen externalista gondolatmenet lehet például az, hogy tudjuk: kognitív képességeink segítségével jól meg tudunk oldani számos hétköznapi alkalmazkodási feladatot, s tudjuk azt is: természettudományos tudásunk segítségével számos olyan célt is el tudunk érni (például a repülés, a születésszabályozás vagy bizonyos betegségek gyógyítása), amelyet a mi természettudományos ismereteinkkel nem rendelkező őseink hiába próbáltak. Mindkét jelenségnek az a legjobb magyarázata, hogy a kívülágra vonatkozó hiteink, illetve az ezeket nyújtó kognitív képességeink általában vagy nagyon sok esetben megbízhatóak.

Meggyőződésem szerint azonban az eddig tárgyalt kétféle válasz csak akkor fenntartható, hogyha P(R/E&N) meghatározhatatlan. Amennyiben ez a helyzet, akkor az emberi megismerés kialakulásának és működésének naturalista leírása (a fejlődéelmélet és a megismerés kialakulásának tudományos – biológiai-pszichológiai – magyarázata) egyszerűen *közömbös* az ismereteink megbízhatóságára vonatkozó kérdést illetően, sem alá nem támasztja, sem nem cáfolja ezt a megbízhatóságot.

<sup>12</sup> Van Cleve 2002. 103. Van Cleve rekonstrukciója a „kartezianus kör”-ről nem biztos, hogy helyes értelmezése a *történeti* Descartes gondolatmenetének, de ez a kérdés a jelen tanulmány mondanivalóját nem érinti.



Más a helyzet akkor, ha  $P(R/E\&N)$  egyértelműen *alacsony*. Egy állat, egy gyermek vagy akár egy reflexiókra képtelen felnőtt kétségkívül tudja (vagy legalábbis igazoltan hiszi), hogy „a hó fehér”, azért, mert úgy látja, hogy a hó fehér – anélkül, hogy bármikor gondolna rá, milyen az emberi érzékszervekre és a környezetre, illetve az emberi megismerésre vonatkozó megfontolások miatt tekintjük a látását megbízhatónak. Ugyanígy egy matematikusnak sem kell matematikafilozófiai problémákon törnie a fejét ahhoz, hogy megbízzon egy bizonyítás érvényességében. De ha a látásban vagy a matematikai bizonyításban ösztönösen bízó személynek elmagyarázzuk, és igazoljuk, hogy egy adott esetben érzékszervei *nem* működnek megfelelően, vagy a matematikai bizonyítás alkalmazott módszere *nem* érvényes, akkor már nem hiszik igazoltan (még kevésbé tudják), hogy a látás vagy a bizonyítási módszer megbízható.

Meggyőződésem szerint tehát ha Plantinga érvelése kifejezetten azt igazolná, hogy a kognitív képességeinkkel megszerzett hiteink valószínűsége csekély ( $P(R/E\&N) \ll 0,5$ ), akkor érvelésével szemben a fentebb leírt két „stratégia” hatástalan.

Ahogy Hume-nak az indukcióra vonatkozó szkeptikus kérdésfeltevése hatására sem tudunk változtatni azon, hogy bízunk az indukcióban, és használjuk azt, ugyanúgy Plantinga érvelése sem tud persze változtatni azon, hogy alkalmazzuk „normál” módszereinket a valóság megismerésére. De ettől még – akár csak Hume esetében – a filozófiai reflexió perceiben el kell ismernünk, hogy az érvelés igenis kihúzza az alapot ismereteink megbízhatósága alól. Ezért nevezi Plantinga a maga érvelését „hume-i megcáfoló megfontolásnak” („*Humean defeater*”).

### III. „KARTEZIÁNUS” VÁLASZOK

Plantinga szerint a naturalizmus alapján az emberi megismerés rendszere *inkoherens*, önmagát cáfolja. Ha pedig ez így van, akkor nem adhatunk olyan választ Plantingának, amely valamely meghatározott ismeretünkre támaszkodik – ahogyan a globális szkeptikus érvekre általánosságban nem adhatunk választ semmilyen ismeretünkre hivatkozva, hiszen ezzel megkerülnénk a kérdést a szkeptikus ellenében. Ha Plantingának az emberi megismerés tényleges kialakulásáról adott rekonstrukciója helytálló, és ha a megismerés megbízhatóságára vonatkozó kérdéseinket *csak* a megismerés mint biológiai-természeti jelenség és képesség eredetére vonatkozó ismereteink alapján válaszolhatjuk meg – akkor az, hogy  $P(R/N\&E)$  alacsony, azt jelenti, hogy  $P(R)$  is alacsony.

De kimutatta-e Plantinga, hogy  $R$  nem kaphat alátámasztást valahonnan *máshonnan*? Kimutatta-e, hogy ismereteink rendszere egészének megalapozása *csak* az emberi lény kognitív képességeinek természetes eredetére, evolúciójára vonatkozó ismereteink révén lehetséges? Ehhez Plantingának azt kellene

igazolnia, hogy a naturalizált episztemológia az *egyetlen* lehetséges episztemológia. Ámde bármennyire is elterjedt a naturalista episztemológia, bármily sokan állítják Sellars, Quine, Rorty és mások nyomán, hogy az ismeretelmélet hagyományos kérdései (mindenekelőtt az episztemológiai szkeptikus kérdésvetése) megválaszolhatatlanok, s ezért nem kell velük foglalkoznunk – még mindig vannak filozófusok, akik továbbra is foglalkoznak a karteziánus értelemben vett episztemológiával – sőt, olyanok is, akik éppenséggel napjainkban *találnak viszsza* hozzá.<sup>13</sup>

Emellett pedig az EAAN a szkepszis egy *sajátos formája*, s így teljesen elképzelhető, hogy az EAAN-generálta szkepszist akkor is meg tudjuk cáfolni, ha az episztemológiai szkepszis más formáit nem.

Az első ilyen „karteziánus” válasz szerint az emberi megismerés rendszerének egészéről *koherensebb* képet kapunk akkor, ha elvetjük az EAAN-t, és megpróbáljuk megmagyarázni, hogyan alakulhatott ki ismereteink megbízhatatlanságának a *látszata*, mint ha elfogadjuk. William Talbott például „a megcáfoltatás illúziójáról” beszél Plantinga érvével kapcsolatban. Ezen azt érti, hogy Plantinga episztemológiai standardjai túl szigorúak. Plantinga szerint pusztán azért, mert ismereteim, világképem kialakulása során *egyszer* (bármilyen indokból) bizonyos kételyek vetődtek fel bennem kognitív képességeim megbízhatóságával kapcsolatban, kétségbe kell vonnom ismereteim egész rendszerét. Talbott a következő analógiákat hozta fel ennek cáfolatára:

Térjünk vissza Plantinga gyógyszeres példájára! Képzeld el, hogy valóban bevettem az X gyógyszert, ezt tudom is, és ezért nem bízom egyetlen kognitív képességemben sem. De ezután felhív telefonon az orvosom, és közli, hogy elvégzett rajtam nemrégén egy genetikai vizsgálatot, amelyből az derül ki, hogy a népességnek abba a 10%-ába tartozom, amely genetikai okokból védett a gyógyszer hatásával szemben. Ekkor – legtöbbször a racionalitás természetéről alkotott elképzelései szerint legalábbis – minden okunk megvan arra, hogy megbízunk az orvos állításában – s így abban is, hogy jól érzékelttem, mikor csengett a telefon, jól hallottam, amit az orvos mondott, s ezért minden más kognitív képességemben is, amelyeket az első megcáfoló megfontolás, a gyógyszer beszédére vonatkozó tudásom kétségessé tett. Az orvos telefonhívása és az általa közvetített információk – valamint az ebből levont következtetések – egy, a saját kognitív képességeim alakulására vonatkozó, a gyógyszerhasználatból következő „forgatókönyvvel” ellentétes „forgatókönyvet” adtak a kezembe.

Talán ez az eset nem megfelelő ellenpélda, mert nem *minden ember* hiteiről, hanem csak a *saját* empirikus hiteim megbízhatatlanságáról van szó. Ezért Talbott konstruált egy másik példát is. Képzeld el, hogy én egy húszas éveim elején járó fiatal ember vagyok, aki szerzett egy könyvet az emberiség történetéről.

<sup>13</sup> Lásd például egy ilyen kísérletként: BonJour 2002.

A könyvet olvasgatva megtudom, hogy akkortájt, amikor anyám terhes volt velem, az emberiség táplálékláncába beépült egy anyag, nevezzük HS-nek, amely minden magzatot súlyosan károsított, olyan módon, hogy születésük után hallucinációkra hajlamos skizofrénekként nőttek fel, akiknek empirikus hitei megbízhatatlanok. Értelemszerűen megrémülök, és érthető módon azt gondolom, hogy az *én* összes kognitív képességeim és hiteim is megbízhatatlanok. Tovább olvasva a könyvet azonban megtudom, hogy egy meghatározott, ritka baktérium ellenanyagokat fejleszt ki a szervezetben a HS-hatás ellen, s az Egészségügyi Világszervezet gyors intézkedése nyomán a baktériumból kifejlesztették a HS megfelelő ellenanyagát, amelyet azonnal kiosztottak a világ népességének – ilyen módon megakadályozva a skizofrénia tömeges elterjedését a következő nemzedékben. Nyilvánvalóan ebben az esetben is racionálisabb azt gondolnom, hogy – bár születésem tájékán nagyon komoly esélye volt annak, hogy megbízhatatlan kognitív képességeim fognak kialakulni – újabb ismereteim fényében valószínű, hogy ez nem így történt (Talbot 2002. 159–160).

Plantinga szerint ez nem így van. Az első esetben ugyanis az orvos telefonhívása (az általa közvetített információkkal együtt) *éppúgy* lehet a gyógyszer indukálta fantázia, mint összes többi, a gyógyszer beszedése után kialakult meggyőződésem – s ugyanez vonatkozik a második esetben a könyv továbbolvasása során szerzett „megnyugtató” információkra (Plantinga 2002. 227–228).

Plantingának igaza van abban, hogy ez a lehetőség *logikai lehetőségként* fennáll. De Talbot szerint felmerül a kérdés: Vajon ha X gyógyszer hatott ránk, és megbénította kognitív képességeinket, akkor miért éppen *ezeket* az illúziókat, miért éppen ilyen, a saját megtévesztő hatását látszólag koherensen leíró s – a mi konkrét esetünkben – meg is cáfoló hiteket alakított ki bennünk? Úgy tűnik, tapasztalatainknak egyszerűbb és koherensebb leírása az, ha elfogadjuk: az orvos valóban telefonált, és az általa elmondottak valóban megcáfolják a kognitív képességeinkre vonatkozó kételyeinket. (Természetesen nem hinnénk az orvosnak akkor, ha azt is tudnánk, hogy az X gyógyszer *pontosan azt* az illúziót alakítja ki bennünk, hogy elveszítettük a kognitív képességeinkbe vetett bizalmunkat, de telefonált az orvosunk, és így tovább... Lásd Talbot 2002. 171.)

Miért ne volna tehát lehetséges, hogy az EAAN által nyújtott, megismerésünk eredetére és működés módjára vonatkozó – szkeptikus végkicsengésű – „forgatókönyvvel” is szembeszegezzünk egy ilyen ellentétes előjelű „forgatókönyvet”, amely úgy írja le kognitív képességeink eredetét és működését, hogy egyrészt végső soron megbízhatónak mutatja őket, másrészt megmagyarázza, hogy mikor, milyen tényezők következtében nem megbízhatóak, vagy tűnnek ilyeneknek? Ahhoz, hogy az EAAN-t preferáljuk az emberi megismerés ilyen leírásával szemben, Plantingának nemcsak azt kellene bizonyítania, hogy kognitív képességeink általában (evolúciós eredetükből kifolyólag) megbízhatatlanok, hanem azt is, hogy miért *éppen ezeket az* – egymással sokféle módon összefüggő, s egymást erősítő – tévedéseket alakították ki bennünk.

Meggyőződésem szerint a Talbott-típusú ellenvetések csak akkor állják meg a helyüket, ha az emberi megismerés eredetéről és működéséről alkotott képük *egyértelműen* koherensebb annál, amit az EAAN nyújt. Sajnos azonban ez nem egyértelmű. Úgy vélem, Talbott analógiái legalább három ponton hibásak.

Az első az, hogy a Taylor által felvetett szkeptikus lehetőségek (a gyógyszer bevétele, illetve a hallucinációkat eredményező skizofrénia tömeges elterjedése az emberiségben) nem igazi *globális* szkeptikus érvek, az EAAN viszont az. Ugyanis a gyógyszeres példában az a lehetőség merül fel, hogy az *én* kognitív képességeim *egy bizonyos időponttól* megbízhatatlanok, a tömeges skizofrénia esetében pedig az, hogy *minden ember* kognitív képességei azok – de ismét csak egy bizonyos időponttól fogva. Ezzel szemben az EAAN *minden ember minden* kognitív képességéről szól az egész emberi történelem folyamán. Így az EAAN-generálta szkepszis egyetemesebb érvényű, mint mindkét Taylor által felvetett gondolati lehetőség.

Az analógiákkal kapcsolatos másik probléma abban áll, hogy semmiféle okunk sincs azt gondolni, hogy akár a gyógyszer, akár a hallucinogén skizofrénia olyan téves meggyőződéseket eredményezne, amelyek bármilyen módon koherensek lennének akár egymással, akár a gyógyszer bevétele, illetve a skizofrénia kialakulása előtt (vagy hatásuk elmúlása után) elfogadott hiteinkkel. Nem így áll a dolog az evolúcióval, s az evolúció Plantinga által elképzelt „valóságtorzító” hatásaival. Természetesen itt nem beszélhetünk arról, hogy milyen hiteink voltak az evolúció „előtt” vagy „után”. Viszont van némi racionális indokunk azt feltételezni, hogy amennyiben az evolúció téves (vagy legnagyobb mértékben téves) hiteket generál bennünk, akkor ezek nem lesznek egymással radikálisan inkohereusak. Az emberiség ugyanis egy többé-kevésbé változatlan természeti feltételrendszerben alakult ki és fejlődött az elmúlt néhány százezer évben. Ha és amennyiben az evolúció megbízhatatlan kognitív képességeket (és nagyjában-egészében téves hiteket) alakított ki bennünk, amelyek azonban mégis hasznosnak bizonyultak a természetes környezetünkhöz való alkalmazkodás során, akkor van némi valószínűsége annak, hogy ezek az illúziók, illetve tévedések legalábbis *koherensek*, bizonyos mértékig összefüggenek egymással, s alátámasztják egymást. Valamennyien ismerünk olyan embereket, akiknek merev, bigott vallási (vagy politikai) meggyőződései vannak, amelyeket semmi sem tud kikezdeni, s ezek a meggyőződések valamiféle „tartást” adnak nekik, sőt sikerekhez vezetnek az alkalmazkodásban. A vallásszociológusok kutatásai tudományosan is igazolják ezt. Olyasmiről viszont nem tudunk, hogy (a drogfogyasztó vagy a skizofrén hallucinációihoz hasonló) állandóan és teljesen inkohereusen változó meggyőződések bármiféle „tartást”, önbizalmat vagy optimizmust adnának bárkinek is.

Végül létezik egy harmadik probléma is Talbott analógiáival kapcsolatban, ez pedig az, hogy a gyógyszeres és a skizofrénias „sztori” esetében is teljesen világos, hogy miért vetődtek fel a kognitív képességeimre vonatkozó kételyek. Ha kognitív képességeim megbízhatóak, és hiteim nagy része igaz, akkor tel-

jesen érthető, hogy bizonyos körülmények között felmerülnek bennem a Talbott által elképzelt lehetőségek, s kételkedni kezdek ismereteim rendszerében. Ezzel szemben az EAAN esetében nem világos, hogy miként és miért alakult ki a „megcáfoltatás illúziója”. Nem világos, hogy ha az emberek kognitív képességei általában megbízhatóak, hiteik általában igazak, akkor miért kezdtek az emberek ebben egyáltalán kételkedni? Nem állítom azt, hogy nem lehetséges a kétely kialakulására olyan racionális magyarázatot adni, amilyent arra adunk: miért gondolnánk a Talbott által elképzelt esetekben, hogy kognitív képességeink megbízhatatlanok. Csupán azt állítom, hogy Talbott gondolatmenete *nem* ad ilyen magyarázatot.

Meggyőződésem szerint azonban meggyőzőbb „kartezianus” válasz is lehetséges az EAAN-ra. Ez a válasz abból indul ki, hogy miközben az EAAN talán igazolja, hogy ismereteink *valószínűleg* tévesek, másfelől bizonyos hiteinkről tudjuk, tudhatjuk, hogy *biztosan* igazak.

Térjünk vissza Descartes példájára! Tételezzünk fel egy groteszk lehetőséget. Azt tudniillik, hogy Descartes *nem* tudott volna eljutni az *idea clara et distincta* vizsgálatából Isten *veracitas Dei*jének igazolásához. Ehelyett Descartes egy olyan Isten hipotéziséhez jutott volna el, akinek a természetéből az adódik, hogy az általa teremtett megismerő alanyoknak az összes hitei (vagy az elméjükben lévő *idea clara et distincta* vizsgálata révén nyert hitei) mintegy 10%-ban igazak. Vajon mindez indokot adott volna-e Descartes-nak arra, hogy *ekvesse* azokat az *idea clara et distincta*ból nyert hiteit, amelyekből rendszere kiindult? Nyilván nem. Ugyanis Descartes 1-nek tekintette ezeknek a hiteknek a valószínűségét, biztosan igaznak tekintette őket, márpedig ha valamiről valamilyen alapon abszolút bizonyossággal tudjuk, hogy igaz, akkor nem jó érv ellene az a megfontolás, amely szerint kilencvenszázalékos valószínűséggel hamis. De annyi kétségtelen, hogy Descartes meglehetősen nehéz érvelési szituációba jutott volna ebben az esetben. További filozófiai hipotézisekkel kellett volna kiegészítenie saját, kognitív képességeinkről adott leírását, olyan hipotézisekkel, amelyek megmagyarázzák, hogy ha bizonyos esetekben az *idea clara et distincta* vizsgálata a megismerés megbízható kritériuma, akkor milyen releváns különbség van ezen esetek és a többi eset, a „megbízhatatlanság” esetei között, továbbá megmagyarázzák azt is, Isten miért úgy teremtett minket, hogy kognitív képességeink csak ezekben a specifikus esetekben legyenek megbízhatóak?

Vajon rendelkezésére áll-e ez a lehetőség a naturalista filozófusnak? Vannak-e olyan hitei, amelyekről biztosan tudja, hogy igazak? Nincs kizárva, hogy igen. Számos filozófus véli úgy – Quine-nak az a priori fogalmán gyakorolt kritikája ellenére is –, hogy bizonyos hiteink önmagukban hordják igaz voltuk vagy legalábbis nagyfokú valószínűségük garanciáját. Ilyenek a filozófiában hagyományosan biztosnak tekintett matematikai, logikai hiteink (amelyek igaz voltát végső soron az ellentmondás elve garantálja), illetve a saját képzeteinkre vagy a saját hiteink tartalmára vonatkozó (introspekció révén nyert) hiteink.

Ha pedig biztosan tudjuk, hogy ezek a hiteink igazak, akkor mindenképpen fenntartásuk mellett kell döntenünk, hiszen az EAAN még a legrosszabb esetben, legszélsőségesebb értelmezése mellett sem azt bizonyítja, hogy *bizonyosan* tévesek, legfeljebb azt, hogy *valószínűleg* azok. Ha viszont ezeket a hiteinket fenntartjuk, akkor – akárcsak Descartes tette – absztrahálhatunk belőlük egy vagy több – legalábbis *prima facie* megbízhatónak tűnő – igazságkritériumot, illetve az e kritériumnak az alkalmazására épülő kognitív képességet. Descartes nyomán gondolhatjuk például azt, hogy azok a hiteink, amelyeket deduktív módon, illetve saját fogalmaink elemzésével nyerünk, valamint amelyek saját tudatunk tartalmaira vonatkoznak, általában igazak. Ezen az alapon felépíthetünk egy olyan ismeretelméletet, amely kognitív képességeinket általában, vagy legalább azok *egyes típusait* megbízhatóknak mutatja, megbízhatóbbnak, mint az EAAN. Innen azután tovább is léphetünk: felmerül a kérdés, hogy ha bizonyos kognitív képességeink mindenképpen megbízhatóak, akkor ez nem teszi-e valószínűvé, hogy a *többi* ilyen képességünk (például az észlelés, az emlékezet és az induktív következtetés képessége) szintén az?

Azok az ismereteink, amelyeket biztosnak tekinthetünk, olyan biztosnak, hogy még az EAAN által felvázolt „szkeptikus forгатókönyv” igaz voltát elismerve sem kell feladnunk őket, talán csakugyan *csak az a priori* meggyőződések, illetve a saját tudatunkra vonatkozó hitek. Ámde felmerül a kérdés: valószínű-e az, hogy az evolúció kizárólag éppen a saját tudatunkra, illetve bizonyos formális, a priori igazságokra vonatkozóan „látott el” bennünket megbízható kognitív képességekkel, illetve igaz hitekkel, miközben szűkebb és tágabb környezetünket illetően viszont nem? Ez a lehetőség teljesen valószínűtlennek tűnik. A természetes kiválasztódás szempontjából fennmaradásunkhoz sokkal inkább szükséges az, hogy szűkebb fizikai környezetünket meg tudjuk ismerni, mint hogy saját tudattartalmainkat/hiteinket vagy bizonyos absztrakt-logikai (esetleg matematikai) igazságokat ismerjünk. Mert még ha igaza is van Plantingának abban, hogy az igazság birtoklása vagy kognitív képességeink megbízhatósága nem feltétlenül jelenti a természethez való alkalmazkodás hasznos eszközét, akkor is nyilvánvalóan a természetes környezetünk megismerésére való képességünk *sokkal inkább* jelent ilyen eszközt, mint a saját tudatunk tartalmainak vagy az absztrakt-formális igazságoknak az ismerete.

Tovább erősíti ennek a megfontolásnak a valószínűségét, hogy apriorisztikus kognitív képességeink természetesen nem a többi kognitív képességeinktől „fállal elválasztva” léteznek az elmében, hanem azokkal szorosan összefonódva is használjuk őket, hiszen a hétköznapi és természettudományos megismerés során folyamatosan alkalmazzuk legalábbis az elemi logikai készségeket (de a matematikát is...) tudásunk ellenőrzésére és rendszerbe szervezésére. Megint csak nem tűnik nagyon valószínűnek, hogy az evolúció olyan – megbízhatatlan – kognitív képességeket fejlesztett volna ki bennünk, amelyeket a megbízható kognitív képességek szerveznek egységes, koherens (ámde nagyjában-egészé-

ben téves) rendszerré. Az a valószínű tehát, hogy ha az EAAN által megcáfolni nem tudott hiteinkben megbízhatunk, akkor *a fortiori* bízhatunk azokban is, amelyeket cáfolni látszik – legalábbis azoknak az ismereteknek az esetében, amelyek érintik saját hétköznapi életünket, a természeti és társadalmi környezetünkhöz való alkalmazkodásunkat.

Amíg tehát Descartes csakugyan nem került volna túlságosan könnyű helyzetbe, ha az *idea clara et distincta* elemzéséből kiindulva egy „10 százalékban megbízható” Isten hipotéziséhez jutott volna el, addig a naturalista filozófus jóval könnyebb helyzetben lesz. Ha egyszer az önmagunk tudattartalmaira vonatkozó, valamint a priori hiteink igazak, s az ezekre vonatkozó kognitív képességeink megbízhatóan működnek, akkor nagyon valószínűtlen, hogy az evolúció úgy alakította volna ki az emberi elmét, hogy *csak* ezek a kognitív képességeink legyenek megbízhatóak, miközben az ezekkel összefonódva s ezekre is támaszkodva működtetett összes többi kognitív mechanizmusunk teljesen megbízhatatlan. (Ugyanakkor e mechanizmusok valamiért mégis olyanok, hogy koherens rendszert alkotnak, megalapozzák és kiegészítik egymást – és csak a filozófiai reflexió perceiben vesszük észre, hogy eredményeik valójában tévesek.)

Itt természetesen csak felvázolni tudtam egy, az EAAN-nal szembeni lehetséges „kartezianus” stratégia körvonalait. Ámde fontos látnunk, hogy a naturalizmus még e stratégia egyértelmű sikere esetén sem „szabadul meg” az EAAN jelentette filozófiai kihívástól. Plantinga ugyanis visszahúzódhat egy második „védővonalra”. Azt állíthatja, hogy bár az EAAN nem bizonyítja azt, hogy a kognitív képességeink megbízhatóságába vetett hit inkonzisztens N&E-vel, de annyit igen, hogy ez a hitünk mindenesetre *konzisztensebb* a teizmussal. Mert hogyha az EAAN-nal szembeni érvelésünk *nem* a naturalista episztemológián, hanem olyan kartezianus, a priori megfontolásokon nyugszik, amelyeket a fentebbiekben kifejtteni próbáltam, akkor a naturalisták nem a naturalizmusból következő indokok miatt bizakodnak saját kognitív képességeik megbízhatóságában, hanem egy olyan nem-naturalisztikus ismeretelméletre hivatkozva, amelynek a naturalizmussal való összeegyeztethetősége legalábbis problematikus. Hiszen egy nem-naturalizált episztemológia a naturalizmus ontológiájába legalábbis nehezen beilleszthető. Viszont a (zsidó-keresztény) teizmus alapján nagyon is kézenfekvő, plauzibilis, hogy nem *csak* az evolúció során kialakult, környezetünkhöz alkalmazkodó biológiai lények, hanem megismerő szubjektumok is vagyunk. P(R/E&N) tehát talán nem olyan alacsony, ámde mindenesetre P(R/E&T) (T-vel jelöljük a teizmust) jóval *magasabb!* Másutt megkíséreltem ezzel az állásponttal is foglalkozni (Szalai, 2009. 1475–1476), ám a jelen tanulmány keretei ezt nem teszik lehetővé.

## IRODALOM

- Beilby, James (szerk.) 2002. *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument Against Naturalism*. Ithaca, Cornell University Press.
- Bergmann, Michael 2002. Commonsense Naturalism. In Beilby 2002. 61–91.
- BonJour, Laurence 2002. *Epistemology: Classic Problems and Contemporary Responses*. Lanham, MD, Rowman & Littlefield.
- Churchland, Patricia S. 1987. Epistemology in the Age of Neuroscience. *Journal of Philosophy*. 84/10. 544–553.
- Fodor, Jerry 2002. Is Science Biologically Possible? In Beilby 2002. 30–42.
- Nunley, Troy 2005. A defense of Alvin Plantinga's Evolutionary Argument against Naturalism. <http://edt.missouri.edu/Winter2005/Dissertation/NunleyT-060605-D1612/short.pdf>
- Plantinga, Alvin 1994. Naturalism Defeated. <http://philofreligion.homestead.com/files/als-paper.htm>
- Plantinga, Alvin 2002. Reply to Beilby and his Cohorts. In Beilby 2002. 204–276.
- Reed, Thomas McHugh 1997. Evolutionary Skepticism. *International Journal for Philosophy of Religion*. 42/2. 79–96.
- Reid, Thomas 1969. *Essays on the Intellectual Powers*. Cambridge, MIT Press.
- Sosa, Ernest 1999. Philosophical Skepticism and Epistemic Circularity. In Ted A. Warfield – Keith DeRose (eds.) *Skepticism. A Contemporary reader*. Oxford, Oxford University Press. 93–115.
- Sosa, Ernest 2002. Plantinga's Evolutionary Meditations. In Beilby 2002. 91–102.
- Szalai Miklós 2005. Fideizmus és posztkartezianus filozófia. *Világosság*. 12. 14–24.
- Szalai Miklós 2009. Darwin kételye. Összeegyeztethető-e az evolúció a naturalizmussal? *Magyar Tudomány*. 12. 1464–1477.
- Talbott, William 2002. The Illusion of Defeat. In Beilby 2002. 153–164.
- Van Cleve, James 2002. Can Atheists Know Anything? In Beilby 2002. 103–128.



# Körlevél az MFT tagjaihoz

Tisztelt MFT tagok!

Amint azt korábbi körlevelünkben jeleztük, az MFT elnöksége és a Magyar Filozófiai Szemle újjáalakult szerkesztősége közötti megállapodás értelmében a Magyar Filozófiai Szemle évi rendes 3 számát az MFT (tagdíjat fizető) tagjai meg fogják kapni. Bár az egyezség szerint csak azok a tagok jutnak hozzá a Szemle számaihoz, akik rendezik tagdíjukat, a Szemle szerkesztőségének jóvoltából a 2010/1. számot most mégis minden MFT tag megkapja. Korábbi körlevelünk mellékleteként minden tagtársunkhoz eljuttattuk az ez évi tagdíj befizetésére szolgáló csekket. Tekintettel a Szemle szerkesztőségével kötött megállapodásra, illetve arra, hogy a Társaság működése a pénzügyi források hiányában meglehetősen nehézkes, az elnökség nyomatékosan kéri a tisztelt tagtársakat, hogy aki még nem fizette be tagdíját, az csekken, vagy a 10102086-09700602-00000002 számú MFT számlaszámra történő utalással rendezze hátralékát. (A tagdíj 2 000 Ft/év.) A Szemle 2010/2. számát ugyanis már valóban csak a tagdíjat fizető tagok fogják kapni.

Az MFT esedékes tisztújító közgyűlésére május 14-én (pénteken) 13.00-kor Szegeden a SZAB Székházban (Somogyi B. u. 7.) kerül sor a *Lábjegyzetek Platónhoz: Az akarat* című konferencia keretei között. (A május 13-14-én megrendezésre kerülő konferencia a Társaság idei nagyrendezvénye.) Az elnökség a tisztújítás előkészítésére Jelölőbizottságot hozott létre. A *Jelölőbizottság tagjai: Mester Béla (mester@webmail.phil-inst.hu), Valastyán Tamás (valastyantamas@freemail.hu) és Pató Attila (patoattila@yahoo.com)*. Az MFT tagjai javaslataikkal és észrevételeikkel a Jelölőbizottsághoz fordulhatnak. A közgyűlésen a leköszönő elnök és főtítkár beszámolójának elfogadását követően kerül megválasztásra a Társaság új elnöke, főtítkára, 2 alelnöke, s további 9 elnökségi tagja, valamint a 3 fős Érdekvédelmi Bizottság. A közgyűléssel, illetve a konferencia programjával kapcsolatos további információk az MFT honlapján lesznek elérhetőek ([www.mft-hps.hu](http://www.mft-hps.hu)). A közgyűlésre és a konferenciára az MFT minden tagját várja a leköszönő elnökség.

Mint azt előző körlevelünkben jeleztük, az MFT Hírek című belső információs kiadvány ettől az évtől nyomtatott formában már nem jelenik meg, így az MFT vezetősége ezután elektronikus úton kívánja tartani a kapcsolatot a tagsággal. Ennek érdekében arra kérünk mindenkit, hogy iratkozzon fel a MaFLa nevű filozófiai levelezőlistára. (Feliratkozás: a [majordomo@hps.elte.hu](mailto:majordomo@hps.elte.hu) címre küldött *subscribe mafla* üzenettel. A feliratkozás után a levelezőlistára szánt üzeneteket a [mafla@phil.elte.hu](mailto:mafla@phil.elte.hu) címre kell küldeni.) Az MFT honlapja ([www.mft-hps.hu](http://www.mft-hps.hu)) ezután is rendszeresen tájékoztatást fog adni a Társaság működéséről, kiadványairól, rendezvényeiről és a hazai filozófiai élet eseményeiről.

Szeged–Budapest, 2010. 04. 12.

Tisztelettel,

az MFT elnöksége



## Naturalizmus a pszichológiában?

Mire is vonatkozik a kérdőjel? Ha szó szerint, mint kifejezést tekintjük, a naturalizmus nem áll különösebben előtérben a mai pszichológiában. Ugyanakkor intellektuális programként mindaz, amit a mai filozófia naturalizmusként emleget, manapság igencsak jellemzi a pszichológia jelentős részét.

Először a pszichológiai naturalizmus kérdésességéről mutatok be néhány adatot. A PsycINFO, a legnagyobb mai nemzetközi elektronikus referáló adatbázis (az 1800-as évektől 29 nyelven mintegy 1 millió szakcikk 38 millió hivatkozással) teljes anyagában 1290-szer fordul elő a *naturalism* kifejezés (először 1896-ban), címben pedig 134-szer. Összehasonlításként: a *materialism* 3824-szer fordul elő, címben pedig 358-szor. Már a számok is mutatják, hova akarok eljutni: miközben a naturalizmus mint kategória s mint terminus nemigen jelenik meg a mai pszichológiában, az a materialista program, amely ennek egyik értelmezése, igencsak előtérben van. Nem állítom, hogy nyeresre áll, mert szociális konstrukcionista barátaim csalást emlegetnének, de jól áll, mint az egyik átfogó felfogás.

Amikor a pszichológiai irodalomban mégis naturalizmust emlegetnek, azt többnyire már a 19. század végi pragmatista értelmezésektől kezdve a vallás, az erkölcs és a nevelés kontextusában teszik. Vajon a természettudományos szemlélet felváltja-e a vallást, megszűnik-e Isten s a teremtés kérdése a tudomány diadalmenete révén? Vajon egyetemes s egységes-e az erkölcs, s vannak-e ennek természeti, értsd biológiai alapjai? S vajon gyermekeink nevelésében csak a konvenciók irányítanak-e, vagy a gyermeki adottságok, a gyermeki természet? Ezek maig fontos kérdések, nem akarom bagatellizálni őket. Brinkmann (2009) például bemutatja, hogy a modern pszichológia mennyire nem marasztalható el egy naiv naturalista tévedésben, hiszen nem tényekből vezeti le az értékeket, hanem az értékek viselkedésirányító szerepét is tényként vizsgálja.

Bishop (2009) egy a vallással, Istennel és a naturalizmussal kapcsolatos vitájában megmutatja, hogy a (természet)tudományos módszer értelmében felfogott és az ontológiailag értelmezett természetelvűség nem ugyanaz a vallást illetően: az elsővel operálva még a kettős igazsághoz is eljuthatunk. Dawkins (2006) és

Dennett (2006), amikor elragadja őket az ateista hév, nem veszik észre, hogy abba az illúzióba esnek, mintha a módszertani naturalizmus egyben ontológiai naturalizmushoz is vezetne abban az értelemben, hogy kiiktatja Istent a világból. Ennek a naturalizmus- és kettősigazság-felfogásnak van magyar vonatkozása is. A magyar–francia katolikus teológus és filozófus Gánóczy Sándor (2008) is ezt az utat követi. Szerinte mind a Biblia helyes, részletes értelmezése, mind a mai neurobiológia arra tanít, hogy az ember mint test és lélek egyetlen lény. Ennek az értelmezésnek megvan a maga teológiai súlya. Gánóczy szerint a dualizmus a keresztény gondolkodás zsákutcája volt. Az emberi test szentsége éppen test és lélek azonosságából érthető meg. Az érvelésnek van világi oldala is. Az egység és a komplex, egészséges működés előtérbe állításával Gánóczy abba a történeti hagyományba illeszkedik a filozófiai pszichológiában, amelyet arisztotelészi funkcionalizmusként emlegetünk (Pléh 2009a).

Mindezt azért (is) érdemes megemlíteni, mert a – Bishop szóhasználatában – módszertani és ontológiai naturalizmus kettőssége a mai pszichológia egészére nézve is felmerül. A kísérleti tudomány elkötelezettjeiként legtöbbször módszertani naturalisták vagyunk, ugyanakkor az ontológiai naturalizmusra nézve szemérmesek. Magam azok közé sorolom, akik nem félnek ontológiai naturalistáknak vallani magukat. Ezt azonban számos klasszikus filozófiai kritika naiv hozzáállásnak tartaná, többnyire azzal marasztalva el – miként Ricoeur érve a Changeux-vel folytatott vita során (Changeux–Ricoeur 2000) –, hogy az élményvilág lefordítása már maga is torzít a kísérleti pszichológiában. Vagyis a jelen szóhasználatban ezek a kritikák a módszertani naturalizmust hibáztatnák.

## I. AZ EMBERI TERMÉSZET ÉRTELMEZÉSE

A pszichológiában tehát a naturalizmus egy tágabb felfogását, ha nem is képviselem, de iránytűként használom, egy olyan felfogást, amely nem az Isten-kérdéssel s az erkölcsi relativizmus problémájával foglalkozik. Felfogásom szerint létezik egy rejtett, tágan értelmezett pszichológiai naturalizmus, amely, akárcsak a modern biológiai naturalizmus (Csányi 1999), az *emberi természet* egy olyan fogalmának a rehabilitálásából indul ki, amelyet mind a marxisták, mind pedig a különböző 20. századi fenomenológiai és társas szabályelvű megközelítések számúztak.

Amint azt az említett adatbázis is mutatja, filozófiai perspektívából ez a pszichológiai naturalizmus Quine-ből (1969, 1999) él. A Quine által meghonosított *naturalized* kifejezés használata *naturalizmus*éval verseng a pszichológiában (544 előfordulás), s Quine is 180-szor fordul elő vagy a „naturalizmussal”, vagy a „naturalizált” kifejezéssel együtt.

Másutt (Pléh 2003) próbáltam részletesen érvelni amellet, hogy mit is jelent ez a naturalista elkötelezettség. Listaszerűen felsorolom egy teljes könyv érrendszerét:

- hit abban, hogy az emberi elme természeti szerveződés, s ilyenként is érthető meg;
- hit abban, hogy az elme mint sajátos, egyéni biológiai szerveződés értelmezendő;
- hit a módszertani individualizmusban, a néplélekszerű képződmények elvetése.

Ez az utóbbi mozzanat, amelyet egy nemzedékkel ezelőtt Fodor (1980) vezetett be egészen nyíltan a kognitív kutatásba (egyenesen módszertani szolipszizmusnak nevezte a követendő felfogást), igencsak vitathatónak tűnik, különösen ma, amikor mindenütt a „társas elme” elismerésének és felismerésének jelentőségét emeljük ki. Ám amint Osbeck (2009) elemzése állítja, a módszertani individualizmus rejtetten sok mai alternatív felfogásban is megmard. Még ha kontextualisták vagyunk is, akkor sem hiszünk néplélekben s hasonlókbán.

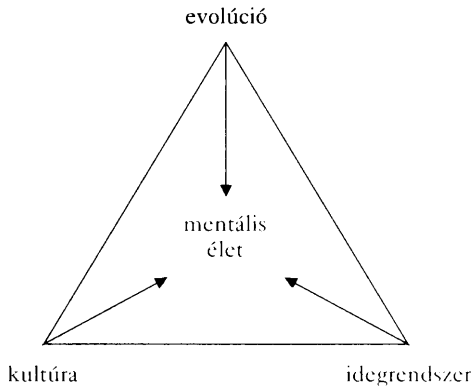
Ha szabad a régi „dialmatos” szövegekre emlékeztetni, az így felfogott naturalizmus szemérmes materializmus. Mi igyekszünk megvédeni ezt a szemérmességet. Emögött ugyanis az rejlik, hogy ez a naturalizmus nem akar előre kész megoldások mellett állást foglalni például test-lélek viszonyában vagy a tudat neurobiológiai értelmezésében. Nem dönti el előre, hogy egyszer majd valamilyen adott agykérgi helyhez vagy éppen egy sajátos működési mintázathoz kapcsolja például az élmény tudatosságát. Csupán annyit hirdet, hogy mindez a tudomány folyamatában mint (természet)tudományos kérdés fog értelmezést nyerni. Ebben az értelemben igazi riválisai azok a tágran értelmezett konstrukcionista felfogások, amelyek az elmét, a gondolkodást diszkurzívan értelmezik, mint egy sajátos beszédmódot, s a tudományt legfeljebb egy beszédmódnak tartják (Bodor 2002).

Ezt a naturalizmust ugyancsak egy nemzedék óta érik a bírálókat. Rom Harré (1989) a mai kognitív tudomány alternatíváival foglalkozva fejtette ki, hogy egy mai konstrukcionista pszichológus számára mindaz elégtelen a naturalista világképben, ami elégtelen volt már Dilthey (magyarul 2004) számára is. Vagyis a naturalizmus nem tud(na) mit kezdeni a mozgó, történeti, interakcióban alakuló kategóriákkal.

A mai pszichológiai naturalizmus esélyét az adja, hogy e kritikákra reagálva ma már nem többszörösen kétpólusú világban, hanem négy pólusban gondolkodik: a lelki jelenséget az agyi folyamatok, az evolúció és a kultúra hármasságában tekinti. Ennek során a konstrukcionista által piederstálra emelt kultúra nem úgy jelenik meg, mint ami megkérdőjelezi a természeti kereteket, hanem annak sajátos részeként. Az optimista naturalista pszichológusok és antropológusok, mint Dan Sperber (2000, 2001), úgy vélik, hogy a pszichológia éppen azért marad meg tudományként, mert agy és kultúra között a pszichológia sajátos témáját képező reprezentációk világa teremt kétirányú kapcsolatot. Sperber könyvének eredeti alcíme is – *a naturalistic approach* – olyan természeti megközelítést

ad, amely bevallottan (nem szemérmesen) kettősen materialista (Sperber 2000). A *mentális reprezentációk* az egyének fejében neurológiai mintázatokként léteznek. A módszertani individualizmusnak megfelelően szerinte ez a mintázatelvű, egyéni materializmus a kiindulópont. (Gondolatok csak fejekben keletkeznek.) Ugyanakkor a mentális reprezentációk nyilvános reprezentációkként – mint anyagi jelek – közléssé válnak: ez a „másik materializmus” Sperber felfogásban. A kultúrákutatók, az antropológia maga, a kétféle anyagi közeg közötti átjárások törvényszerűségeinek mintázati tudománya.

A kortárs pszichológiai naturalizmusnak az új vonása abban áll, hogy ma a pszichológusok egyszerre veszik figyelembe a proximális és a disztális mozzanatot: Magát a szocialitást is az emberi természet részének tekintik, s egy sajátos, új „szemiotikai háromszögben” képzelik el a működési terepét. Ezt sok konstrukcionista is elfogadná, ők azonban a belső nyilatkat nem oksági viszonyokként, hanem „interpretációkként” kezelnék (Moghaddam–Lee–Harré 2007).



A biológiában használt kétféle hozzáállást, a proximális (idegrendszeri) és a disztális (evolúciós) magyarázatot a mai kutatók egyszerre tekintik. Azt mondják például – egy egyszerű esetet véve –, hogy a színek perceptuális világának megértéséhez két biológiai értelmezésre van szükségünk: mi valósítja meg a színek élményrendszerét az idegrendszerben (ez a proximális oldal), s miért jött létre, mi volt az adaptív haszna a színlátásnak (ez a disztális oldal)? Természetesen itt nem csupán kiegészítő „foltozgatásról” van szó, hanem arról, hogy a két attitűd magyarázatként egyszerre feltételezi és korlátozza egymást.

Idegrendszer, evolúció és mentális folyamatok kapcsolatának vizsgálatában a mai naturalista pszichológiának fontos új jellemzője az, hogy a társasság elemi formáit is az emberi természethez sorolja (lásd Pléh 2008). Eközben – amikor például a prefrontális területek, az érzelmi leképezés és a szociális felelősség kapcsolatát vagy az én-reprezentációk keletkezését vizsgálja (Damasio 1996, Changeux–Ricoeur 2001) – sokkal kidolgozottabb szociális koncepciókat használ.

Ebben a felfogásban a kultúra mint evolúciósan kialakult beállítórendszer jelenik meg, nem pedig mint olyan tényező, amely mintegy kivezetne a természetből. A környezeti interakció, amit hagyományosan tanításnak nevezünk, valójában sajtós szelekciós ágensként működne. Ennek nagy befolyású modellje Tomasello (2002) felfogása vagy Gergely György és Csibra Gergely modellje a pedagógiai hozzáállásról (Gergely 2002, Csibra–Gergely 2007). Ezek a tanító kultúra mint figyelemirányító rendszer tanulásszerkezetet is meghatározó jelentőségét emelik ki. Egy jellegzetes példáját ennek a kultúra-természet illesztésnek Stanislaw Dehaene mutatta meg az írás rendszerére vonatkoztatva. Agyi képalkotó módszerekkel azt találták, hogy a legkülönbözőbb írásrendszerek esetén lényegében ugyanott: a parieto-occipito-temporális területen, vagyis a fali, a tarkó- és a halántéklebeny találkozási helyén alakul ki a vizuális szóforma terület. Ezt a területet, amely a finom vizuális feldolgozásra alakult ki evolúciósan, mintegy újraverbuválja az írás mint kulturális rendszer. Dehaene és Cohen ezt mondják erről:

A neurális újra-ciklálás a következő feltevésekből indul ki. 1. Az emberi agyi szerveződést az evolúciótól örökölt erős anatómiai és kapcsolati korlátok jellemzik. A csecsemőkorban már igen korán megjelennek a rendezett neurális térképek, és ezek a későbbi tanulást bizonyos irányba befolyásolják. 2. A kulturális elsajátításnak (például az olvasás elsajátításának) meg kell találnia a maga neurális fülkáját, vagyis olyan idegrendszeri hálózatokat, amelyek elég közel állnak a kívánt működéshez, és elég hajlékonyak ahhoz, hogy neurális erőforrásaik egy jó részét erre az új használatra irányítsák át. 3. Amikor az evolúciósan végbement működésekre dedikált kérgi területeket új kulturális tárgyak szállják meg, ezek korábbi szerveződése nem teljesen törlődik ki. Vagyis a korábbi neurális korlátok jelentős hatást gyakorolnak a kulturális elsajátításra és a felnőttkori szerveződésre. (Dehaene–Cohen 2007. 384–385.)

Egy dolgozatomban ennek kapcsán a következőképpen próbálok spekulálni:

Izgalmas kérdés, hogy nem lehet-e maga a nyelv is ilyen? Nem lehetséges-e, hogy a természetes nyelv sem úgy alakult ki, hogy a Broca- és Wernicke-terület speciálisan adaptívan szelektálódott a nyelvre, hanem valamilyen más funkcióra választódott ki, mint például a gyorsan változó mozgások produkálására (ez lenne a Broca-terület), illetve a gyorsan változó hangminták elemzésére (Wernicke-terület)? A nyelv, amikor létrejött mint emberek közötti gyakorlat, ugyanúgy, mint az írás, meglévő agyi struktúrákat verbuvált volna. Dehaene metaforája erre a neurális újra-ciklálás: mikor a kulturális fejlődés felhasznál meglévő idegrendszeri struktúrákat, akkor nem biztos, hogy magára a kulturális újításra, például a nyelvre is kell legyen egy adaptációs történetünk. (Pléh 2010b. 12.)

## II. A DETERMINIZMUS KÉRDÉSE ÉS A PSZICHOGENETIKA

Természetesen régi probléma – visszavezethető Arisztotelész és Platón szembeállításáig, valamint az újkori racionalista–empirista vitáig –, hogy milyen viszony van a genetikai mozzanatok és a környezeti interakciók között az emberi fejlődés meghatározásában. Ez a klasszikus „természet *versus* nevelés” dilemma a mai természettudomány fejlődése következtében a korábbiaknál sokkal komplexebbé vált, ugyanakkor ez az a kérdéskör, ahol a naturalista pszichológia személete a legszorosabban érintkezik az ember determináltságának kérdésével. Vajon amikor azt hirdetjük, hogy sajátos megismerési és érzésmódjaink az emberi természet tág értelmezésében genetikailag meghatározottak, egyben egy fatális, lemondó determinizmust képviselünk-e?

A válasz a genetika diadala közepette a kérdésre: nem. Az embert nem fenyegeti lineáris genetikai determináció. Tézisszerűen fel lehet sorolni, hogy melyek azok a tényezők, amelyek révén a modern genetika eredményei az egész szervezet, és ezen belül a mentális rendszer egyéni kibontakozásának, valóban fejlődésének komplex felfogását állítják előtérbe (Pléh 2010b).

a) *A genom korlátai.* A mai genetikának a pszichológiát is érintő fontos felismerése, hogy az emberi genom sokkal korlátosabb, mint azt korábban gondoltuk, kevesebb mint 30 ezer génnel számolhatunk a fejlődés meghatározásában. Ez nagyon nehezen vetíthető le azokra a felfogásokra, amelyek számos egyedi viselkedési vonást, például olyan nyelvi tényezőket, mint mondjuk az érzékenység a nyelvtani egyeztetésre, bizonyos ízek iránti preferencia stb. mind genetikai meghatározottságúnak tartanak. Szakítanunk kell a viselkedésgenetikában és a pszichológiai folyamatok genetikai értelmezésében is az egy az egyben való megfeleltetés egy olyan eszményével, amely minden viselkedési fenotípusnak egy egyedi gént feleltetett volna meg. A legtöbb viselkedési mozzanat genetikai értelemben poligénes meghatározottságúvá válik.

b) *Szabályozó gének előtérbe állítása.* Testi tulajdonságok esetében is számos igen részletesen kidolgozott és bizonyított genetikai modell van arra nézve, hogy a gének sajátos időzítő, szabályozó genetikai rendszerek hatálya alatt fejtik ki hatásukat. Azt, hogy például egy állat testfelülete milyen szelvényezettségű és csíkozottságú lesz, nem a szelvényezettségért felelős külön gének, hanem bizonyos sejtszaporodási folyamatokat irányító időzítő, szabályozó gének befolyásolják. Ezek a szabályozó gének rendkívül jelentősek az idegrendszer fejlődésében és következésképpen az emberi fejlődésben oly fontos szerepet játszó kritikus periódusok meghatározásában is.

c) *A nyitott rendszer genetikája.* Az emberre leginkább jellemző mentális folyamatok, például a nyelv, nyilvánvalóan nyitott rendszerek abban az értelemben, hogy az egyéni élet során történő beállítást feltételeznek, de ugyanez érvényes az arcpreferencia-rendszerre és így tovább. Olyan genetikai modelleket kell tehát keresnünk a fejlődés meghatározóinak vizsgálatában, ahol magában



a genetikai modellben mintegy elvárásként benne van a környezeti interakciók világa, amelynek megfelelően stabilizálódik például már az első év során az anyanyelv hangrendszere egy sokkal nagyobb repertoárból és így tovább (Mehler–Dupoux–Gervain 2008).

Vagyis a fejlődés a genetikai értelmezésben is egy sajátos út, amelyen pályák, domborzatok és néha – a genetikai eredetű fejlődési zavarokban – szakadékok vannak, de nem valamiféle predetermináció.

Míndez tehát nem azt sugallja, hogy a mai naturalizmus kész királyi út lenne a pszichológiában. Ugyanakkor határozott kutatási programot körvonalazó út, mely az új nemzedéket is képes orientálni. Nem eliminálja, hanem természeti folyamatként magyarázza a mentális folyamatokat.

## IRODALOM

- Bishop, Robert C. 2009. What is This Naturalism Stuff All About? *Journal of Theoretical and Philosophical Psychology*. 29. 108–113.
- Bodor Péter 2002. Konstrukcionizmus a pszichológiában. *BUKSZ*. 14. 67–74.
- Brinkmann, S. 2009. Facts, Values, and the Naturalistic Fallacy in Psychology. *New Ideas in Psychology*. 27. 1–17.
- Changeux, Jean-Pierre 2008a. *Az igazságkereső ember*. Budapest, Gondolat Kiadó.
- Changeux, Jean-Pierre – Ricoeur, Paul 2000. *A természet és a szabályok*. Budapest, Osiris.
- Changeux, Jean-Pierre 2008b. *Du vrai, du beau, du bien. Une nouvelle approche neuronale*. Paris, Odile Jacob 2008.
- Csányi Vilmos 1999. *Az emberi természet*. Budapest, Vince.
- Csibra G. – Gergely Gy. (szerk.) 2007. *Ember és kultúra*. Budapest, Akadémiai Kiadó.
- Damasio, Antonio R. 1996. *Descartes tévedése*. Ford. Pléh Csaba. Budapest, AduPrint.
- Dawkins, Richard 2006. *The God Delusion*. Boston, Houghton Mifflin.
- Dehaene, S. – Cohen, L. 2007. Cultural Recycling of Cortical Maps. *Neuron*. 56. 384–398.
- Dennett, Daniel 2006. *Breaking the Spell. Religion as a Natural Phenomenon*. New York, Viking.
- Dennett, Daniel 2008. *Darwin veszélyes ideája*. Ford. Kampis György – Kavetzky Péter. Budapest, Typotex.
- Dilthey, Wilhelm 2004. *A történelmi világ felépítése a szellemtudományokban*. Ford. Erdélyi Ágnes. Budapest, Gondolat.
- Fodor, Jerry 1980. Methodological Aolipsism Considered as a Research Strategy in Cognitive Psychology. *Behavioral and Brain Sciences*. 3. 63–110.
- Ganoczy, A. 2008. *Christianisme et neurosciences*. Paris, Odile Jacob.
- Gergely György 2002. Az intencionális cselekvések értelmezése és utánzása csecsemőkorban. In Czigler I. – Halász L. – Marton L. M. (szerk.) *Az általánostól a különöségig*. Budapest, Gondolat.
- Harkai Schiller Pál 1940/2002. *A lélektan feladata*. Budapest, Osiris.
- Harré, R. 1986. Social Sources of Mental Content and Order. In J. Margolis – P. T. Manicas – R. Harré – P. F. Secord (eds.) *Psychology: Designing the discipline*. London, Blackwell. 91–127.
- Harré, R. 1988. Wittgenstein and Artificial Intelligence. *Philosophical Psychology*. 1. 105–115.

- Harré, R. 1989. Vygotsky and Artificial Intelligence: What Could Cognitive Psychology Possibly be About? *Philosophical Psychology*. 2. 389–400.
- Mehler, J. – E. Dupoux – J. Gervain. 2008. *Ember születik*. Ford. Gervain Judit. Budapest, Gondolat.
- Moghaddam, F. N. – Lee, N. – Harré, R. 2007. Psychology is Social. Exploring Universals in Performance Capacity and Performance Style. *Rivista di Psicologia*. 25. 139–176.
- Moore, G. E. 1981. *Principia Ethica*. (részletek) Ford. Lónyai Mária. In Lónyai Mária (szerk.) *Tények és értékek*. Budapest, Gondolat. 51–105.
- Osbeck, L. M. 2009. Transformations in Cognitive Science: Implications and Issues Posed. *Journal of Theoretical and Philosophical Psychology*. 29. 16–33.
- Pléh Csaba 2003. *A természet és a lélek*. Budapest, Osiris.
- Pléh Csaba 2008. *A pszichológia örök témái*. Budapest, Gondolat.
- Pléh, Csaba 2009a. *History and Theories of the Mind*. Budapest, Akadémiai.
- Pléh Csaba 2009b. A kognitív tudomány flörtje és civakodós házasságai a filozófiával: Wundt köpenyéből bújnának ki Fodor moduljai? *Világosság*. 50/3. 55–65.
- Pléh Csaba 2010a. A pszichogenetika kutatási stratégiái és egy példa: A Williams szindróma. *Magyar Tudomány*. 170. 49–57.
- Pléh Csaba 2010b. Darwin és a pszichológusok. *Iskolakultúra*. 2. 2–21
- Quine, W. V. 1969/1999. Naturalizált ismeretelmélet. In Forrai Gábor – Szegedi Péter (szerk.) *Tudományfilozófia*. Budapest, Áron Kiadó. 369–382.
- Quine, W. V. 1969. *Ontological Relativity and Other Essays*. New York, Columbia University Press.
- Sperber, D. 2000. Metarepresentations in an Evolutionary Perspective. In D. Sperber (szerk.) *Metarepresentations*. New York, Oxford University Press.
- Sperber, D. 2001. *A kultúra magyarázata*. Budapest, Osiris.
- Waddington, Conrad Hal 1957. *The Strategy of the Genes*. London, George Allen – Unwin.

## A tudomány heterogenitása és a naturalizmus\*

### I. BEVEZETÉS

A naturalizmus megjelenése a filozófiában hosszú történeti folyamat eredménye. A logikai empirista (ismertebb nevén logikai pozitivistá) hagyomány már a századelőn „naturalista” módon állt a filozófiai kérdések megválaszolásához, de az „ősök” keresése könnyedén folytatható lenne még korábbi fejlemények, úgy mint Mach, Du Bois-Reymond és a 19. századi tudományos filozófia említésével.<sup>1</sup> A Quine nevével fémjelzett hagyomány is erről a töről fakadt (bár az elődök kapcsán elmehetünk akár John Dewey-ig), és tipikus követői a filozófusi háttérrel rendelkező, de a tudományos kérdések iránt nyitott amerikai értelmiségiek voltak, a logikai pozitívizmus második és harmadik generációjának képviselői. Ugyanakkor az álláspontok részben már korábbi filozófiai elköteleződésekhez kapcsolódtak, például a naturalizmus ontológiai tézise egy, a materializmus szofisztikálódott fizikalista változatára hajaz. Ráadásul a naturalista szemlélet megjelent a filozófiai hagyományon kívül is, amikor természettudományok képviselői próbáltak számot adni hagyományosan filozófiai kérdésekről, (nem meglepően) egyetértve a naturalista filozófusokkal a (természet)tudomány episztémikus felsőbbrendűségének tézisében. A tág értelemben vett naturalista megközelítések közös jellemvonása, hogy a hagyományosan a filozófia területére tartozó kérdések legalább egy részénél (triviálisan bizonyos episztemológiai problémák megoldásánál, de akár etikai kérdésekben is) szükségesnek tartják a releváns tudományos válaszok ismeretét és felhasználását.

Ez a nagyon tág jellemzés számos megközelítési irány lehetőségét tartalmazza, és ezek kibontakozásának nyomán évtizedekig tartó, tanulságos filozófiai vita

\* A szerzők a tanulmányt Bolyai János posztdoktori ösztöndíjasként írták, a munkát támogatta az OTKA K 72598, a Magyar Állami Eötvös Ösztöndíj és a Collegium Budapest. A szerzők köszönetet mondanak Demeter Tamás megjegyzéseire.

<sup>1</sup> Rövid történeti ismertetéshez lásd Rheinberger 2010. Az előtörténethez tartozik a normatív és deskriptív dimenziók elválasztásának igénye is, amelynek erős szociológiai háttere is volt (Kusch 1995); a „Hogyan jutunk el vélekedéseinkhez?”, illetve a „Hogyan kellene eljutnunk vélekedéseinkhez?” típusú kérdésekre más-más diszciplínák keresték a választ, ez esetben a pszichológia és a filozófia (Kornblith 1985).

zajlott. A naturalista álláspontokat képviselők változó mértékben hagyják jóvá, illetve írják át a diszciplináris határokat, részben annak függvényében, hogy a normatív kérdések megválaszolásában milyen szerepet szánnak a deskriptív kérdésekre adott válaszoknak (egyes körökben pl. mantraszerű az „ought implies can” ismétlése). Vannak egészen ideológiai színezetű naturalista álláspontok, amelyek a filozófiai álláspontok felhasználásával átfogó világgépeket hirdetnek, és itt ugyanúgy megtaláljuk a merev determinizmust naturalizmussal legitimálókat, mint a panteizmus védelmezőit.<sup>2</sup>

Ez a tanulmány elsősorban a filozófiai naturalizmus episztemológiai kérdéseit vizsgáló hagyományt elemzi, miközben szem előtt tartja, hogy ezek a leszűkítések és határmegvonások problematikusak (és részben szociológiai alapon tehetőek meg, pl. a publikációk helyeinek meghatározásával), nem egyértelműek és kikezdzhetők. Fő tézisünk az, hogy a tudományok elemzése és empirikus vizsgálata az utóbbi évtizedekben komolyan megkérdőjelezte annak a tudományképnek a tarthatóságát, amely a filozófiai naturalizmus legtöbb típusának hátterében áll. Ez pedig azt jelenti, hogy a naturalista megközelítsek egy jelentős része, miközben a tudományos válaszok primátusát hirdeti, éppen a saját előfeltevéseinek tudományos megalapozatlanságát nem ismeri fel.

Az ezt alátámasztó érvek megvizsgálásához azonban néhány előzetes szempont elfogadása szükséges. Az egyik az, hogy a filozófiai naturalizmust nem olyan restriktív értelemben használjuk, ahogy azt Quine tette, aki számára az „új ismeretelmélet [...] a pszichológia egy fejezeteként része a természettudománynak” (Quine 1999, 377). Egyrészt a Quine-t követő hagyomány szakított a szigorú alárendelési tézissel, és a filozófusok jelentős része nem fogadja el a filozófia autonómiájának teljes feladását és a filozófia eliminálhatóságát – az érvekhez lásd pl. (Almeder 1993). Kortárs cikkekben a filozófiailag sokkal gyengébb tézist (amely egyben a filozófiának sokkal nagyobb mozgásteret ad) szokták érvelés nélkül elfogadni, mely szerint a tudományról *bizonyos kérdések kapcsán informálnak* kell lennie a filozófusnak. Másrészt nem tűnik *a priori* eldönthetőnek, hogy mely megközelítés lehet alapja a naturalizálásnak; jól ismert, hogy Quine választása a behaviorista pszichológiára esett, és jól ismert ennek a kutatási hagyománynak a kudarca is. Vannak, akik naturalista szellemben – az evolúcióelméletet segítségül hívva – amellet érvelnek, hogy ez *a posteriori* ügyis eldől, éppúgy, ahogy a tudományok esetében eldől az, hogy melyik módszer vagy értelmezési keret bizonyul gyümölcsözőnek (Giere 1985; Giere 1988). Ezt erősíti az, hogy a – módszertani – naturalizmust elfogadók egyetértenek abban, hogy az elméleti racionalitásnak nincsenek *sui generis* normái, az elméletválasztás

<sup>2</sup> Tom Clark munkája (Clark 2007) jól jellemzi a naturalizmust világgépként értelmező és az utóbbi években egyre erősödő áramlatokat. Ezek az irányzatok részben a tudománnyal kapcsolatos társadalmi viták hatására alakultak ki, mint pl. az evolúció-kreáció vita, ahol filozófusok is komoly szerepet vállaltak, mint Alvin Plantinga.

során a racionalitási normák *következményei* számítanak, vagyis az általuk megszerzhető igaz hitek alapján ítélnéjük meg őket (Feldman 2006).

Ha a naturalizálás sokféle lehet, akkor *a priori* nem kizárható olyan tudományterületek ismereteinek figyelembevétele, mint a tudományt tudományosan vizsgáló kognitív tudomány, tudománytörténet vagy tudomány- és tudásszociológia. Ezek alapvetően empirikus elemzéseket adó deskriptív vállalkozások, amelyek különböző mértékben kvantifikáltak ugyan, de módszertanaik összemérhetőek a tudományos módszerekkel, és szociális struktúráik és gyakorlataik megfelelnek a tudományos intézményekkel, eljárásokkal szemben támasztott elvárásoknak.

A következőkben azt próbáljuk vázlatosan megvizsgálni, hogy ha episztémikusan kitüntetettnek tartjuk a tudományos tudást, és ezt a bizalmat nem korlátozzuk sem a behaviorista pszichológiára, sem a természettudományok bármelyik szeletére, hanem kiterjesztjük minden empirikusan alátámasztott és ma tudományosnak tekintett tudásterületre, akkor a tudományt vizsgáló tudományok mai tudása alapján a filozófiai naturalizmusról és annak előfeltevéseiről mit mondhatunk. A kiindulási problémánk az, hogy első megközelítésben inkoherensnek véljük az alábbi három állítás együttesét:

- a) A tudománytanulmányok megállapítása szerint a tudomány lényegileg heterogén vállalkozás.
- b) A tudománytanulmányok a tudományt empirikusan vizsgáló tudományterület.
- c) A naturalista elismeri episztémikus függését a tudománytól.

Az inkoherencia abban áll, hogy amennyiben a naturalista hajlamos elfogadni igazként a tudományos válaszokat, ugyanakkor elismeri azt, hogy a tudomány-nyal kapcsolatos tudományos válaszok a tudománytanulmányoktól származnak, úgy tehát ez utóbbi terület válaszait is el kell fogadnia, ezek között azt is, hogy a tudomány nem ad egyöntetű válaszokat – így tehát nem létezik „a” tudomány, amitől episztémikusan függene. Ezért az a megalapozási probléma, amit a filozófiai naturalizmus részben feloldani, részben megkerülni próbált, új módon (és a résztvevők által gyakran nem elismerten és felismerten) jelenik meg.

Annak érdekében, hogy naturalistaként megpróbáljuk feloldani e látszólagos ellentmondást, sorra vizsgáljuk a fenti három állítást és következményeiket. A 2. szakaszban azt vesszük szemügyre, mit mondanak a tudománytanulmányok a tudomány egységességének kérdéséről. A 3. szakaszban tárgyaljuk azt a problémát, hogy el kell-e fogadni a tudománytanulmányok megállapításait mint tudományos leírásokat, vagyis hallgatnia kell-e a naturalistának az elmúlt évtizedek empirikus tudományvizsgálataira. A 4. szakasz azt a kérdést veti fel, hogy a heterogén tudomány koncepciója milyen tanulságokkal jár az episztémikus függési tézis értelmében a naturalizmus ontológiai felfogásai szempontjából, az 5. szakasz pedig ugyanezt a kérdést vizsgálja a módszertani

naturalista értelmezések esetén. A 6. szakasz levonja a tanulságokat, és megkísérel feloldani a látszólagos inkohereciát a naturalizmus fogalmának egy tartatóbb értelmezésével.

## II. A TUDOMÁNY SOKFÉLESÉGE

A hagyományos logikai pozitivista megközelítés nem kérdőjelezte meg a tudományos módszer unikalitását, céljának tekintette univerzálisnak hitt igazolási és magyarázati sémák részleteinek kidolgozását a tudomány különböző területeire.<sup>3</sup> A tudományt empirikusan elemző történészek és szociológusok azonban ezt az egységes alapot nem találták meg a különböző tudományok vizsgálatakor. Ez persze nem jelenti azt, hogy elvileg lehetetlen ennek megtalálása, de a komoly erőfeszítések sorozatos kudarca a tudományfilozófusok jelentős részét (is) arra készítette, hogy elfogadják a tudás-termelési módszerek heterogenitását, és ezeket figyelembe véve dolgozzák ki elképzeléseiket – a tudomány episztémikus kitüntettségének megkérdőjelezése nélkül (Dupré 1993; Galison–Stump 1996; Hacking 1983). Ma a tudományt kutató tudósok általános vélekedése, hogy a tudomány (mint minden országban jelen lévő társadalmi intézmény) számos szinten vizsgálva heterogén, és ez a heterogenitás lényegi tulajdonsága. A tudósok eltérő társadalmi közegekben felnöve mind a világról, mind a megismerésről eltérő nézeteket vallanak, ahogy a Nyugat és a Kelet összehasonlító elemzései is megmutatják (Nisbett 2003). Még egy tudományterületen belül is különböző kulturális hátterek jelennek meg az érvelési mintázatokban.<sup>4</sup> Mindezek mellett a tudománytanulmányok kutatásai hatalmas mennyiségű ismeretet halmoztak fel arról, hogy a hely, a lokalitás jelentősen befolyásolja a tudomány fejlődését és a tudósokat (összefoglalóhoz lásd Henke és Gieryn 2007).

Ennél talán még jelentősebb, hogy történeti fejlődésük során a különböző tudományterületeken eltérő normák alakultak ki azzal kapcsolatban, hogy mi számít igazolt / alátámasztott eredménynek, milyen magyarázati sémák fogadhatók el.<sup>5</sup> Ezek a változások kontingensek, de mind a tágabban vett kulturális hatások, mind a szűkebb szakterületek fejlődése, mind a tudományterületek sajátos

<sup>3</sup> A Bécsi Kör egyik utolsó vállalkozása egy hatalmas munka, az „Egységesített tudomány nemzetközi enciklopédiája” kiadása volt, amely sok (14 vagy 26) olyan kötetet foglalt volna magába, melyek maguk is egyenként 10-10 önálló könyvnyi tanulmányból álltak volna össze (Morris 1960). (Ebből aztán csak az első két kötet valósult meg.) A munka címe által kifejezett elköteleződés természetesen a pozitivista hagyomány öröksége.

<sup>4</sup> Kevin Dunbar például kanadai, olasz és USA-beli immunológiai kutatócsoportok vizsgálata során (2002) azt találta, hogy az amerikai kutatók gyakrabban használnak analógiákat, míg az olasz kutatók (habár egy részük az USA-ban tanult) inkább deduktív sémákkal operálnak.

<sup>5</sup> Ez utóbbinak talán legkitűnőbb példáját kínálják a biológiai magyarázatokkal kapcsolatos viták, ahol a mai kutatói gyakorlat és a történeti szempontok egymást erősítő módon jelennek meg (Bechtel 2006, 2008; Craver 2007).

tárgyi kultúrája és társadalmi szerveződése által részben meghatározottak. Így időszakonként és tudományterületenként változnak a fizikai modellek, a diagramok és ábrák olvasási módjai és értelmezései, de ugyanígy változnak az absztrakt modellek – mint az ideális gáz, harmonikus oszcillátor, döntéseméleti vagy populációgenetikai modellek, lásd (Chadarevian–Hopwood 2004). Mindezek fényében egyáltalán nem meglepő, hogy a magyarázattal szemben támasztott elvárások, az, hogy mit tekinthetünk pl. oknak, hogy vannak-e követelmények a matematikai kifejtéssel szemben, és ha igen, milyenek, hogy mi fogadható el „törvénynek” – mindez egy tudományterületen belül is sokféle lehet.

Az utóbbi évtizedekben nemcsak a képek és egyéb nem diszkurzív reprezentációs formák hatását vizsgálták a tudományban (Lynch–Woolgar 1990), hanem azt is, hogy a kutatói gyakorlatok mi módon befolyásolják a különböző tudományterületeket (Galison 1987, 1997; Golinski 1998; Latour–Woolgar 1986; Rheinberger 1997, 2006). A példák még sorolhatók lennének, hiszen olyan fontos szempontokat, mint a tudományterületek közötti aszimmetrikus tudásátvételt (és episztémikus hierarchiát), az eltérő társadalmi gyakorlatokat (mint bírálati rendszerek különbsége, az eredmények publikálásának stiláris megköötöttségei) még nem is említettünk.<sup>6</sup> De talán ennyi elég is, térjünk rá a következményekre, hiszen ennek a heterogenitásnak az elfogadása különös kihívásokat rejt a naturalizmus számára.

Míg a módszertani homogenitás, az univerzális magyarázati normák feltételezése (legyen az akár deduktív-nomologikus, akár induktív vagy egyéb norma) valószínűsítette az egységes tudomány felé történő haladást és így a naturalizálás különböző változatainak konvergenciáját, egy diverzifikálódó és heterogén tudomány képe azt sejteti, hogy a különféle naturalizáló felfogások radikálisan eltérő állításokat tesznek. Ezt a gyanút táplálja az is, hogy konkrét esettanulmányok elemzésekor még egy diszciplínán belüli elméletválasztás is jelentős különbségeket eredményez, és eltérő normatív állítások támaszthatók ezekkel alá.<sup>7</sup>

Persze egy naturalista számára fennáll az a lehetőség, hogy elfogadja a tudományok heterogenitására vonatkozó megállapítást empirikusan érvényesként, miközben fenntartja a hitét abban, hogy naturalista magyarázatoknak a hetero-

<sup>6</sup> Különösen meggyőző elemzését adja a diszciplínák radikális különbségeinek Knorr-Cetina (1999), aki a nagyenergiás fizika és a molekuláris biológia „episztémikus kultúráit” hasonlítja össze számos szempontból, és tudományantropológiai módszerek segítségével sokrétűen rávilágít a szerveződési és megismerési gyakorlatok alapvető eltéréseire.

<sup>7</sup> A tudományfilozófiai irodalomban erre jó példa a geológiában forradalmi Wegener-féle kőzetlemez-elmélet körüli viták elemzése. A filozófusok egy része konkrét tudományos elméletek felhasználása nélkül írt elemzést (Frankel 1978; Laudan 1980; Thagard–Nowak 1990), ahol azonban a naturalizálás megjelent, ott nagyon eltérő elméletek felhasználásával nagyon eltérő konklúziókat támasztottak alá (Giere 1988; LeGrand 1988; Solomon 1992; Stewart 1990). A kognitív tudomány felhasználása a naturalizált ismeretelméletben pedig egyre több – és nem egyre kevesebb – alternatívát tud felmutatni, ahogy ezt nagyon jól érzékelteti Carruthers, Stich–Siegal 2002.

genitásból származó sokfélesége eliminálható lesz valamikor a jövőben. Ehhez a heterogenitást kell pusztá kontingenciának tekinteni, ami megfelelő ráfordítással és a tudományok fejlődésével megszüntethető. Tarthatjuk tehát megőrzendőnek az egységes tudományba vetett hitet (sokak szerint délibáb), ekkor azonban magyarázatot kell adnunk arra, hogy az episztémikusan kitüntetett tudást létrehozó tudomány jelenleg miért nem ily módon szerveződik. Kereshetünk magyarázatot ennek az egységességnek a hiányára, azonban ez a magyarázat egyben érvként lesz felhasználható amellet, hogy a tudomány, mivel nem követi ezt a gyakorlatot, nem a létező legjobb tudástermelő intézmény, vagyis episztémikus kitüntetettsége megkérdőjelezhető.

Ugyanakkor a heterogenitás felismerése a deskripció szintjén sokakat – részben konzekvens naturalista hozzáállásuk miatt – újfajta normatív állítások, és éppen ellenkező irányú következtetések megfogalmazására ösztökélt. Hiszen ha a tudomány a világ episztémikusan kitüntetett megismerési formája, és ennek jellemzője a heterogenitás, akkor nem kizárható, hogy ez a jellemző hozzá is járul a tudomány episztémikus sikeréhez.<sup>8</sup> A tudomány empirikus kutatásai pedig számos olyan modellt és fogalmat vetettek fel, amelyek a sokféleség és az együttműködési kényszer kettősségéből következő fordítási és cserefolyamatok konstitutív szerepét vizsgálják a tudományos tudás termelésében.<sup>9</sup> Ha tehát valaki ragaszkodik az egységes tudomány idealizált képéhez, és ahhoz a feltevéshez, hogy az ennek megvalósulása felé való haladás garantálja a jelenleg még különböző naturalista megközelítések konvergenciáját, akkor neki azoknak a normatív és deskriptív modelleknek is cáfolatát kell adnia, amelyek éppen a heterogén gyakorlatot tartják episztémikusan kitüntetettnek.

Ráadásul mára erősen megkérdőjeleződött az az előfeltevés, hogy a tudás-szerzés során komoly kognitív erőforrásokat érdemes fordítani arra, hogy a világról szerzett ismereteink (gyakran csak potenciális) inkonzisztenciáit kiszűrjük, és egységes rendszerbe foglaljuk tudásunkat.<sup>10</sup> Episztémikus szempontból akár hasznosabban is tölthetjük időnket – vagyis jobban kihasználhatjuk a kognitív erőforrásokat –, ha az empirikus sikerességet maximalizáljuk, és nem a tudományos munka során kialakuló diszkurzív rendszerek átjárhatóságát, redukálható-

<sup>8</sup> A disszenzus általános méltatása mellett (Sunstein 2003) megjelentek azok a modellek, amelyek részletesen próbálják alátámasztani a konszenzushiány episztémikus kitüntetettségét (Solomon 2001, 2006a, 2006b), magyarul részletesebben lásd (Zemplén 2010). Természetesen ezek sem új fejlemények, a késői Kuhn evolúciós („speciációs”) modelljében is találunk érveket a tudomány heterogenitásának episztémikus előnyeiről (Kuhn 1990, 1993), és Feyerabend filozófiájának szimpatikus rekonstrukciója is ebbe az irányba mutat (Farrell 2003), amely visszakövethető a 19. századig, például Mill filozófiájáig (Lloyd 1997).

<sup>9</sup> Ilyen pl. a határtárgy (Star–Griesemer 1989) vagy a kereskedelmi zóna (Galison 1997). Részletesebb összefoglalásért lásd (Kutrovátz–Láng–Zemplén 2010).

<sup>10</sup> A gondosabb rekonstrukció már a logikai pozitívizmus tudásképét és enciklopédizmusát sem ilyenek mutatja, lásd pl. Neurath nézetéről Cartwright et al. 1996.



ságát stb. tekintjük elsődlegesnek.<sup>11</sup> És ekkor még nem is említettük, milyen megvalósítási korlátokkal kell szembenéznie egy olyan vállalkozásnak, amelyik a sokféle (saját gyakorlataikkal és normáikkal rendelkező) tudományterület egységesítését kísérelné meg.

### III. A TUDOMÁNYTANULMÁNYOK LEGITIMITÁSA

Akik ezekre az empirikus felismerésekre legyintenek – és valljuk be, a filozófusok hajlamosak ezt tenni –, miközben naturalistáknak vallják magukat, tapasztalatunk szerint leggyakrabban arra az álláspontra helyezkednek, hogy fenntartják annak a tudományképnek az elfogadhatóságát, amelyet a lassan két generáció óta szakmaszerűen művelt és előbb idézett tudományterületek kezdtek. Vagyis egy deskriptív célú vállalkozási tudományos eredményeit nem fogadják el, miközben a tudományt kitüntetett episztémikus státuszúnak tekintik.<sup>12</sup> Ez a látszólagos ellentmondás természetesen feloldható, ha részletes indoklással jár együtt, amely elfogadható érveket hoz fel arra, miért – és milyen alapon – nem tekinthetők tudományos eredménynek, vagy legalábbis megalapozott tudományos eredménynek az empirikus tudománykutatások belátásai.<sup>13</sup> Mindez összefüggésben van annak megválaszolásával, hogy milyen feltételek teljesítése esetén tekinthető tudományosnak, illetve nem tudományosnak egy

<sup>11</sup> Itt gondolhatunk azokra a szerzőkre, akik a kognitív munkamegosztás mellett érvelnek (Kitcher, de még inkább Solomon és Longino), de felidézhetjük azokat a fejleményeket is, amelyek a magyarázatokkal szemben támasztott igényeket elemzik, arra a következtetésre jutva, hogy számos elvárás, pl. az unifikálás nem szükséges *desideratuma* a magyarázatoknak (Morrison 2000). Mindez kapcsolható a racionális elméletválasztás körüli vitákhoz, amelyek eredményeképp számos, az elméletek közötti választásban kitüntetettként kezelt érték (mint az egyszerűség) szerepe csökkent, és hangsúlyosabbá vált (egyes esetekben egyedüli mércévé vált) az empirikus sikeresség. Ennek az elvárásnak a szempontjából a heterogenitás nem jelent problémát, sőt adott esetben előnyt is jelenthet, amiből persze nem következik az, hogy *önmagában* lenne érték (mint azt Feyerabend „virágozzék száz virág” jelmondata sejteti).

<sup>12</sup> Természetesen elképzelhető, hogy a kritikus szemlélő kétségbe vonja a tudománytanulmányok deskriptív státuszát, és valamiféle filozófiai megközelítést lát bennük. A tudományosság igénye azonban a szándék szintjén megkérdőjelezhetetlen, hiszen alapvető szerepet játszik például a tudományos tudás szociológiájának ún. erős programjában (Bloor 1992), illetve az e nyomán kibontakozó tudománykutatások számos korai irányzatában. A diszciplína ma is elsősorban empirikus jellegű esettanulmányok mentén halad, tartózkodva a filozófiai elméletek megfogalmazásától. Bár újabban vannak próbálkozások a normatív megfontolások rehabilitációjára (pl. Collins–Evans 2002; Latour 2004; Lynch 2006), ezek részben éppen azért kapnak komoly figyelmet, mert vállalásukban erősen eltérnek a szakma főként implicit konszenzuális elköteleződéseitől. Ha viszont valaki elismeri a deskriptív tudományosság szándékát, de ennek sikerességét vitatja, akkor demarkációs gyakorlatot végez – az ezzel kapcsolatos problémákat lásd a továbbiakban.

<sup>13</sup> Egy ilyen típusú válasz összhangban van azzal az ellenségességgel és esetenként leplezetlen lenézéssel, amellyel a tudományok ismeretelméleti hívei gyakran tekintenek a tudományelemzések különböző (és leginkább filozófiai és társadalomtudományos) formáira. Lásd ehhez az ún. tudományháborút (pl. Gross–Levitt 1994; Sokal–Bricmont 2000).

szakterület (visszavezetve a diskurzust a demarkációs problémához), illetve miért van különböző tudományterületeknek eltérő státuszuk a filozófiai naturalizálás során.

Ez a gyakori válaszstratégia azonban éppen amiatt problematikus, mert amikor a naturalista filozófus diszkreditálja a tudománykutatásokat valamilyen explicit vagy implicit demarkációs kritérium alapján, akkor feltételezi, hogy kompetens egy terület tudományosságának megítélésében. Ennek indoklására három lehetőség tűnik kézenfekvőnek. Az első az, hogy a kérdéses filozófus fenntartja a tudományfilozófia hagyományos igényét a demarkációs kritérium megfogalmazására. A második szerint meghagyja ezt az igényt a tudósoknak, és olyan tudósokra hivatkozik, akik áltudományosnak tekintik az empirikus tudománykutatásokat (lásd az előző lábjegyzetet) – vagyis naturalista elköteleződését kiterjeszti a tudományos kérdésekről a meta-tudományos kérdések felé (amelyeket szintén a tudósok hivatottak megválaszolni). A harmadik megoldás az, ha a tudomány területére utalja ugyan a demarkációs kérdések megválaszolását, de vindikálja magának az ilyen ítéletek megfogalmazásának jogát. Vizsgáljuk sorra a fenti lehetőségeket!

Az első lehetőség a tudományfilozófia történetének fényében meglehetősen bizarrnak tűnik. Bár az ún. demarkációs probléma tényleg komoly figyelmet kapott a terület klasszikus szerzőinél, a tudományfilozófiai konszenzus már egy ideje lemondott az ezzel kapcsolatos igényeiről (Laudan 1983; Vermeir 2006). Ráadásul a demarkáció igénye erősen normatív ambíciókat fejez ki, ami legalábbis nem magától értetődő viszonyban áll a naturalista elköteleződéssel. Ha ugyanis azt tartjuk, hogy a tudomány mondja meg, miként van a világ, akkor különös emellett azt gondolni, hogy a filozófus mondja meg, mi számít tudománynak. Ez az álláspont egy erős meta-episztemikus kontrollt biztosít a filozófusok számára a tudás egésze felett, ami nemcsak ismeretelméleti, hanem etikai és társadalomfilozófiai problémák sokaságát veti fel.<sup>14</sup>

A második lehetőség szerényebb igényekkel lép fel: nem a filozófus kompetens a tudománytanulmányok áltudományként való diszkreditálásában,<sup>15</sup> hanem maguk a tudósok. Feltehető ugyanis, hogy a gyakorló tudósok jobban tudják, mi számít tudománynak, mint a tudományelemzők. Ez az érvelés azonban körkörös, két szempontból is. Egyrészt azért, mert aki a (természet)tudósok ítéletére bízta magát abban a kérdésben, hogy legitim tudomány-e a tudománytanulmányok területe, és ezek alapján arra jut, hogy nem, az episztemikus forrásának megválasztásakor már eleve elfogadta azt az ítéletet, hogy a tudományelemzők *nem* kompetensek ennek eldöntésében, ha ugyanis azok lennének, akkor őket

<sup>14</sup> Természetesen léteznek ilyen kísérletek – magyar áttekintéshez és példához lásd (Laki 2010).

<sup>15</sup> Bár érdekes, hogy éppen a demarkációprobléma elvetőjeként is ismert Larry Laudan filozófusként már korán érvel a kérdéses terület tudományos státusza ellen (Laudan 1999).

kellett volna megkérdeznie. Érdeemes ugyanis különbséget tenni a tudomány megismerő gyakorlata és a tudományról/ szülő megismerési gyakorlat között. Az utóbbi egy meta-diszciplína (mint amilyen a politika számára a politológia, az irodalom számára az irodalomelmélet, stb.), és tárgyában nyilvánvalóan eltér azoktól a gyakorlatoktól, amelyeket elemez. Ha feltételezzük, hogy egy tudományos gyakorlat által előállított tudás megbízhatóságát a tudástermelés feletti intézményes kontrollnak köszönhetjük (ellenőrzési és kritikai gyakorlatok sokasága), akkor minden tudós csak a saját szakterületén tekinthető kompetens kognitív forrásnak, amelyen képezték, publikál, és nap mint nap megmérteti magát. Ebben az esetben a tudományelemzés kérdésében nem az akármilyen gyakorló tudós lesz kompetens, hanem éppen az, amelyiknek a tudományelemzés a szakterülete, és azt legitim szakértőként képviseli.<sup>16</sup> Vagyis amennyiben létezik ez a szakma, és az ettől származó ismeretek gazdagítják a tudományról alkotott tudásunkat, akkor nem elegendő általában a tudósokra hallgatni, hanem az erre specializálódott tudósok véleménye lesz mérvadó – vagyis pont azoké, akit ki akarunk zárni.<sup>17</sup>

A fenti válasz körkörös abból a szempontból is, hogy amikor valaki „a tudósok” ítéletére hagyatkozik a demarkáció tekintetében, akkor azzal az előfeltevéssel él, hogy a tudósok által adott válasz egyöntetű lesz csakúgy, miként a tudomány által kínált egyéb válaszok is azok – márpedig mi azt állítjuk, hogy nem ez a helyzet.<sup>18</sup> Ha például a társadalomtudományhoz fordulunk a demarkációprobléma kapcsán, akkor ott egészen más választ kapunk, mint ami a tudósok szakmailag kevésbé reflektált ítéleteiben általában megszokott: itt pl. ideológiai erőfeszítésként jellemezhető a határmegvonási kísérlet, átutalva ezzel a problémát a normatív dimenzióból a deskriptívbe (Gieryn 2006). A vizsgált válasz tehát megkerüli az őt kiváltó kérdést, amennyiben eleve elfogadja a tudomány egységességének előfeltevését akkor, amikor egyes tudósok demarkációs kritériumait vagy ítéleteit azonosítja „a” tudomány demarkációs kritériumával, és nem veszi figyelembe a disszenzus jelenlétét (vagy akár lehetőségét).

A harmadik válaszlehetőség az volna, hogy a naturalista filozófus tudományos problémának látja ugyan a demarkáció problémáját, amelyben tudósoknak kell

<sup>16</sup> A kérdéshez szorosan tartozik az ún. metaszakértelem fogalma, amely a tudománytanulmányok egyik legnépszerűbb kurrens irányzatának, a „szakértelem-tanulmányoknak” egy lényeges eleme (Collins–Evans 2007), ahol a „szakértelem periódusos rendszere” egyaránt magában foglalja a laikusok tudománnyal kapcsolatos véleményét és a hivatásos metaszakértők ítéleteit. A kérdésről magyarul lásd bővebben Kutrovácz 2010a; 2010b.

<sup>17</sup> Megkíméljük az Olvasót attól, hogy saját területünk legitímációjához fogjunk: mivel az intézményes feltételeket tekintve ezt egy legitim tudásterületnek látjuk, feltesszük, hogy aki ezzel nem ért egyet, az magára veszi a bizonyítás terhét. Mi itt most pusztán amellet érvelünk, hogy a legitimitás megkérdőjelezése milyen problémákat von maga után a cikk gondolatmenete szempontjából.

<sup>18</sup> Sőt, a tudósok ún. metatudományos (vagyis a tudományról alkotott) ítéletei még nagyobb heterogenitást mutatnak, mint szakterületük tárgyszintű állításai, ehhez lásd Shapin 2006.

kompetens ítéletet hozni, ám mégis megtartja magának a jogot arra, hogy ilyen ítéletet fogalmazzon meg. Ahogy az 5. szakaszból ez részletesen ki fog derülni, ez a válasz elmosza a határvonalat tudós és naturalista filozófus között, vagy legalábbis folytonosságban látja a tudományt a filozófiával. Látni fogjuk, hogy mi akár elfogadhatónak is tekintünk egy ilyen megoldást, azt azonban fontos észrevenni, hogy a naturalista alapfelfogás implicite posztulálja tudomány és filozófia világos különbségét akkor, amikor a tudományos válaszokat tartja elismerendőnek a filozófiai vizsgálat számára. Ha ugyanis ezt a megkülönböztetést eltöröljük, akkor a naturalista tézis átfoglalozása („a tudomány egy területe – a naturalista filozófia – a tudomány más területeinek válaszait elfogadja”) kiüresedik, és a tudomány és filozófia viszonyát feszegető kérdést a tudomány kognitív munkamegosztásának problémája váltja fel.

Mindazonáltal nem gondoljuk, hogy a naturalista attitűd és a heterogén tudomány képe elvileg összeegyeztethetetlen volna, de úgy véljük, hogy a fentiekben felvetett nehézségek alapjaiban problematikusá teszik számos naturalista felfogás jogosultságát. A továbbiakban azt vizsgáljuk meg, hogy milyen következményekkel jár a tudomány heterogenitásának tézise a naturalizmus különböző formáira nézve.

#### IV. ONTOLÓGIAI NATURALIZÁLÁS

Az ontológiai tézis a létezőkre vonatkozó egzisztenciális döntéseket a tudósokra bízta. Ha azonban a tudósok nem egységes módszerekkel és egységes magyarázatokkal dolgoznak, könnyen előfordulhat, hogy más létezőket tekintenek valónak. Mit tehetünk ilyenkor? Régi megoldási kísérlet a tudományok hierarchiájának felállítása, amely alapján a fundamentálisnak tekintett diszciplína létezőit tekintjük valónak, míg a többi diszciplína ontológiai kijelentéseit ezekre a létezőkre redukálhatónak tartjuk (Oppenheim–Putnam 1958). A fizikát tekintve ilyen fundamentális tudománynak aggodalomra ad okot, hogy a kutatási gyakorlat szintjén időnként maguk a fizikusok is egyszerre használnak egymással inkonzisztens elméleteket, amelyekkel új és elfogadott tudást állítanak elő (Falkenburg 1996), valamint hogy a fizikán belül sem sikerült a generációk által remélt konszenzust megteremtteni a nagy elméletek között.

A fizikánál maradvány – bár a példa általánosítható – az ontológiai naturalizmus különösen izgalmas problémája, hogy a tudósok és a filozófusok számára más jelentőségű az ontológiai szint. Ez a fizikán belül is jelentős történeti ingadozást mutat a 17. századi mechanikus filozófiától a 18–19. századi newtoniánus fizikán keresztül a mai napig. Míg például Maxwell elektrodinamikai elméletének kifejtésekor elengedhetetlen lépésnek tekintette a mechanikai modell kidolgozását (amely még éterrészcscékkel operált), a mai fizikusok a Maxwell-egyenleteket problémamentesen elfogadják, holott ma már nem te-

kintik az éterrészecskéket létező entitásoknak, és így megkérdőjeleződik, hogy a hullámok miben terjednek. A matematikai modellek empirikus sikere akkor is elfogadandóvá tesz egy elméletet, ha a matematikai apparátusban posztulált entítások ontológiai realitással nem bírnak, sőt már vagy százéves az a hagyomány, amely a tudomány fejlődését elsősorban a matematikai, absztrakt, szimbolikus reprezentációk fejlődésében látja, az egzisztenciális kérdéseket pedig *nem tudományos kérdésnek* tartja (Duhem 1954).<sup>19</sup> Saját tapasztalatunk az, hogy míg a legtöbb tudós inkább instrumentalista, és kevésbé tartja fő céljának az ontológiai kérdések megválaszolását,<sup>20</sup> addig a legtöbb filozófus inkább realista, és vagy feltételezi, hogy a tudósok egyetértenek az ontológiai kérdésekben, vagy ha ismer is vitás eseteket, a tudósoktól várja ezek megválaszolását (ebből az is kiderül, hogy kevés fenomenológust ismerünk).

Még mindig a fizikát véve példának könnyen ütközünk komoly szemantikai kihívásokba is. Az elektron létezését száz éve bizonyítottnak véve megdöbbenő, hogy mennyire más elképzeléseket tartottak igaznak a tudósok erről az entitásról (a pudingban megbúvó mazsolától a szigorúan egy síkban az atommag körül keringő kis részecskén át az állóhullámig, sőt a lokalizálhatatlan valószínűségeloszlással leírható entitásig). A tudomány szemantikai hozzáállását jól leíró finitista szemantika (Barnes–Bloor–Henry 2002) jelentősen eltér a mai filozófiában dominánsnak tekinthető *in*finít szemantikáktól.<sup>21</sup> A referenciával szemben támasztott filozófiai követelményeknek nem felel meg a tudomány gyakorlata, maximum az inkább didaktikai leegyszerűsítéseknek tekinthető tankönyvek.

Ezeknek a problémáknak a kifejtése azonban elvezet az utóbbi évtizedekben riasztó papírtengert generáló realizmusvitához és a filozófia szemantikai elköteleződéseit feszegető problémákhoz, ezért talán a legbölcsebb a rendelkezésre álló terjedelmet másra használni. Térjünk vissza a tudományok hierarchiájának kérdéséhez! Oppenheim és Putnam híres cikke (Oppenheim–Putnam 1958) „munkahipotézisként” fogalmazta meg a tudományok hierarchiájának modelljét, az elmúlt fél évszázad azonban inkább diszkonfirmálta, mintsem megerősítette

<sup>19</sup> A matematika esetében sokkal gyakoribbak az irrealista álláspontok, mint pl. a normatív tények esetében. Ugyanakkor nemcsak az elméleti entítások ontológiai státusza változhat az időben, hanem az is, hogy milyen elterjedt a realista elköteleződés egy tudományterületen belül. A századelőn az instrumentalista pozíció sokkal elterjedtebb volt pl. az elméleti fizikában, mint ma. És fordítva: a matematikában a nem-euklideszi geometriák felfedezését megelőző időszakban a geometriai elmélet alapértelmezésben realista volt.

<sup>20</sup> Az ontológiai elköteleződések kontextusfüggők lehetnek. Reuben Hersh matematika-filozófus gyakran idézett mondása szerint „a tipikus matematikus hétköznap platonista, ünnepnapokon formalista” (Hersh 1979), vagyis realista elköteleződésekkel dolgozik munkája során, ám ha ezekre rákérdezzük, akkor antirealista pozícióba menekül. A gyakorló természettudósok ontológiai elköteleződései általában hasonlóan rugalmasak.

<sup>21</sup> Itt érdemes olyan vitákra gondolni, mint a merev jelölők és a természeti fajták kérdése. Számos nagy karriert befutott filozófiai nézet tarthatóságát radikálisan kikezdi a tudomány empirikus vizsgálatának eredményei (Chang megj. alatt).

ezt a feltevést.<sup>22</sup> A speciális tudományok filozófiájának robbanásszerű fejlődése az elmúlt évtizedekben legalábbis nem tartotta szükségesnek e munkahipotézis elfogadását.<sup>23</sup> Bár a munkahipotézis még rengeteg munkát generál, az olyan típusú problémák megoldása, mint az elme naturalista értelmezése egy fizikalista keretben, számos olyan nehézséggel néz szembe, amelyet a tudomány utóbbi évtizedeinek empirikus vizsgálata tárt fel. A tudományterületek (pontosabban a kutatók azon csoportjai, akik közös problémákról és hasonló módszerekkel szinkron kutatásokat folytatnak) őrzik önállóságukat, ellenállnak az eliminálási kísérleteknek (nem törekszenek például saját konceptuális rendszereik felszámolására és redukálására egy másik fogalmi rendszerben), fejlődésük folyamán korlátozottan lépnek interakcióba más területekkel, a konszenzuálissá váló tudás egy területen csak lassan és korlátozottan jut el más területekre (mire az eredeti területen esetleg már fel is bomlott a konszenzus), stb. Ha csak ezeket emeljük ki az elmúlt évtizedek során szerzett ismeretek közül, akkor is megkérdőjeleződik az ontológiai naturalista program, és az utóbbi évek munkáinak áttekintése ezt megerősíti. A filozófia ezen területén dolgozók létrehozták saját kutatási hagyományukat, ahol az ontológiai kérdéseket rendre olyan előfeltevések mentén fejtik ki, amelyek aligha tükrözik a tudományok valódi gyakorlatát, hiszen a korábbi generációk (inkább ideologikus, mint empirikus alapú) tudományképeinek örökösei.

Mindez nem jelent problémát akkor, ha ezeket a vállalkozásokat hagyományos filozófiai tevékenységnek tekintjük. Ha azonban a vállalkozást naturalizálni állítjuk be, akkor ezt a diszkrepanciát valamilyen módon kezelni kell. Megoldást jelenthet a tudomány empirikus kutatásából származó ismeretek relevanciájának elutasítása a filozófia számára – ez tetten érhető abban a növekvő távolságban, ami az egyre kontextualizáltabb tudománytörténeti és szociológiai munkákat választja el az egyre absztraktabb tudományfilozófiai vizsgálódásoktól (a modell-elméleti realizmus fejlődése egy lehetséges példa az utóbbira). Az előbb felsorolt szociológiai jellemzők mindenesetre alkalmazhatók erre a szakterületre is: a filozófusok őrzik önállóságukat, ellenállnak az eliminálási kísérleteknek, fejlődésük folyamán korlátozottan lépnek interakcióba más területekkel, a konszenzuálissá váló tudás egy területen csak lassan és korlátozottan jut el más területekre (mire az eredeti területen esetleg már fel is bomlott a konszenzus),

<sup>22</sup> Még az a néhány példa, amelyet filozófusok a redukció klasszikus és sikeres példáinak tartottak is megingott a tudománytanulmányok részletesebb elemzései által (Kitcher 1998), átfogó értékeléshez az itt felvázolt redukciós program problémáiról lásd (Brigant 2008). Természetesen ezek a kijelentések vitathatók, és a biológiában ma is nagy és izgalmas téma a redukció kérdése, de általában a biológiai gyakorlat *alapján* próbálják a redukció kritériumait meghatározni. Itt elég azt leszögezni, hogy – talán a termodinamika makroszkopikus és statisztikus megfogalmazásának bizonyos értelmezésén kívül – a tudományokkal empirikusan foglalkozók a hagyományos nageli redukció példáit nem tudták meggyőzően rekonstruálni a tudományok történetének és gyakorlatának vizsgálatával.

<sup>23</sup> Sőt, részben ennek a megközelítésnek kritikájaként jelent meg Fodor 1974.

stb.<sup>24</sup> Bár hősiesnek tekinthetjük az ontológiai naturalizmus részletkérdésein dolgozók munkáját, a jelenlegi ismereteink alapján a vállalkozás legnépszerűbb változatai problematikusak és nehezen értelmezhetőek konstruktív naturalista módon.

Nézzünk erre egy példát! A mai ontológiai naturalizmus egyik legfontosabb tézise a kauzális zártság, vagyis annak feltételezése, hogy fizikai okok keresethők minden fizikai hatás mögött. Tehát ha például írunk egy cikket a megújuló *Magyar Filozófiai Szemle* első számába, akkor ennek *a)* oksági magyarázatát lehet adni, *b)* az okok megadhatók pusztán a fundamentálisnak tartott fizikai erők / hatások és a világ fizikai leírása segítségével. Történetileg a kauzális zártság feltételezése nagyon termékeny munkahipotézisnek bizonyult a kanti-blumenbachi teleomechanikus biológiától a szerves kémia és élettan fejlődésén keresztül számos tudományban. Ugyanakkor ez az időszak jól ismert és radikális változást hozott abban, hogy mit is tarthatunk oknak egy oksági magyarázat során. Ha azonban a filozófusok szerepét tekintjük a folyamatban, akkor a naturalizmus programja szempontjából súlyos nehézségekbe ütközünk. Míg ugyanis számos filozófus próbál érvelni az általa helyesnek tartott álláspont mellett *a)* és *b)* kapcsán, addig a naturalizmusból az következik, hogy a létezéssel kapcsolatos kijelentéseket a tudományra bízunk, így amikor a kauzális zártságot védjük vagy támadjuk vagy értelmezzük, a naturalistának fejet kell hajtania az előtt, hogy a tudósok közössége mit fogad el pl. fundamentális erőnek (ami a filozófus oksági magyarázatában legitim módon megjelenhet).

Míg a természetfilozófusok (és a filozófusok egy része) hagyományosan tettek ontológiai állításokat (pl. feltételezve azt, hogy a világon minden anyagi), és ezek tartalmas állítások voltak, addig a naturalizmus ontológiai változata ezeket a döntéseket átadja a tudósoknak. Így azonban könnyen abba a csapdába esik, hogy nem kínál ontológiailag tartalmas álláspontokat, hiszen annyiban naturalizál, amennyiben egyes kategóriák meghatározását a tudományra bízta – ahol ezt nem teszi, ott nem is tekinthető ontológiai értelemben naturalizálnak a megközelítés. Érdekes dilemmához vezet ez: az ontológiai naturalista feltárhat analitikus kapcsolatokat, megfogalmazhat kondicionálisokat, de ezek mögé nem tehet tartalmas egzisztenciális elköteleződéseket.<sup>25</sup> Ha mégis tesz, akkor a tu-

<sup>24</sup> Ezek a tanulságok persze a szociológiára is igazak lehetnek, és semmi esetre sem kritikaként értendők: ilyen a tudomány mint a legmegbízhatóbb tudást előállító társadalmi intézményünk általában, és ilyen az analitikus filozófia is, amely – számos jellemzőjében – megpróbál tudományossá válni. Megmaradva a deskripció szintjén, ebből semmiféle normatív állítás nem következik automatikusan.

<sup>25</sup> Itt konkrét konfliktusba kerül a módszertani és az ontológiai naturalizmus, ha elfogadjuk Papineau jellemzését: „Röviden, a módszertani naturalista ragaszkodhat ahhoz, hogy bárki, aki »komoly metafizika« iránt érdeklődik, tartalmas egzisztenciális állítások megfogalmazásával kell, hogy kezdjen a népi elméleteinkkel kapcsolatban... Azok a további fogalmi elköteleződések, amelyeket *a priori* megfogalmazhatunk, semmilyen filozófiai jelentőséggel nem bírnak” (Papineau 2007). Fontos megjegyeznünk, hogy a fenti érvelésünk az ontoló-

dományok ontológiai prioritását kérdőjelezi meg (tehát nem naturalista a fenti értelemben), ha pedig nem, akkor nem a világról beszél.

Mindez itt azért különösen érdekes, mert a tudomány ontológiai elköteleződéseinek heterogenitásával és másodlagosságával kapcsolatos korábbi megfontolások azt implikálják, hogy a tudománynak átengedett döntések tartalmas volta megkérdőjeleződik, miközben a naturalista filozófus lemondana a tartalmas állítások megfogalmazásának jogáról. Így tehát érvelhetünk filozófusként a kauzális zárttság mellett, ha okként azt fogadjuk el, amit mások (pl. fizikusok) mondanak, bár a tudósok akár egy adott diszciplínán belül sem értenek egyet abban, mit tekintenek oknak (még ha a matematikai reprezentációban egyet is értenek). Kijelenthetjük, hogy egy teljes (fizikai) elmélet megoldja a metafizikai problémáinkat, még ha nem is tudjuk, mi az, és nem mi döntünk arról, hogy mi legyen – ám az ilyen elmélet megfogalmazásának lehetősége erősen kétséges. Kutathatjuk a mentális tények és a fizikai tények kapcsolatát, ahol késhegyig menő harcokat vívnak a különböző nézetek képviselői, de azt, hogy mi számít fizikai ténynek, nem a filozófus dönti el, a tudós számára viszont a probléma nem ebben a formában vetődik fel. Vagyis a tudósok nem arra a kérdésre adnak választ, amelyre a naturalista filozófus választ remél, így a filozófus „tudományos” ontológiája a tudományos elméletek egy értelmezése marad. Ez az értelmezés viszont aktív folyamat, nem pusztán passzív átvétel, így kérdés, megfelel-e a filozófiát a tudomány alá rendelő ontológiai naturalizmus igényeinek.

## V. MÓDSZERTANI NATURALIZMUS

A naturalizmus azonban nemcsak ontológiai, hanem episztemológiai programként is értelmezhető, sőt történetileg ez megközelítés a jelentősebb: gondoljunk csak Quine szándékára, hogy a megismerés filozófiai elméleteit pszichológiai leírásokkal helyettesítse. Így a naturalista hozzáállás széles körű alkalmazása azt célozza, hogy kiiktasson, de legalábbis nyitottá és fallibilissé tegyen minden olyan filozófiai elmélkedést, amelynek tárgyával kapcsolatban létezik tudományos leírás és magyarázat. Módszertani szempontból ez azt jelenti, hogy a naturalista csak olyan mechanizmusokat, folyamatokat, entitásokat, stb. enged meg a magyarázataiban, amelyekhez a tudományos leírások folyamodnak. Ám felvetődik a kérdés: mi az, amit nem ilyennek tekintünk?

Elméletileg közelítve ez a megszorítás abban a közkeletű véleményben nyilvánul meg, hogy magyarázataink csak természeti folyamatokra hivatkozhatnak,

---

giai naturalizálás problematikusságáról sok ponton kapcsolható Crane és Mellor klasszikus fizikalizmus-bírálatához (Crane–Mellor 1990). Megközelítésünk más pontokról indulva tár fel belső inkonzisztenciákat, általánosabb, és előnye, hogy nem épít néhány olyan elemre, amely mind Crane és Mellor, mind az általuk kritizált fizikalista szerzők tudományképeinek részei.



és kerülniük kell a bármilyen értelemben vett természetfelettre való hivatkozást – maga a „naturalizmus” terminus is erre utal. Ha elfogadjuk (a legtöbb naturalistával egyetemben), hogy a természet megismerésének legitim birtokosai a természettudományok, akkor ahhoz a vélekedéshez jutunk, hogy a természettudományok szolgáltatják a világgal kapcsolatos tartalmas magyarázatok lehetséges eszköztárát. Ez a megfogalmazás azonban erősen korlátos. Míg azokban az időkben, amikor a mai naturalizmus gyökerei visszanyúlnak (lásd cikkünk elején a történeti utalásokat) valóban a természettudományok számítottak *par excellence* tudományoknak, és a szellem- és társadalomtudományok tudományos státusza erősen vitatott volt, addig mára az arányok megváltoztak. A társadalmi és kulturális jelenségek körét ma tudományos eszközökkel vizsgáljuk, és ezek a tudományterületek jelentős intézményes súllyal és kognitív tekintéllyel rendelkeznek (egy választás könnyebben megjósolható, mint az időjárás). Ha tehát a társadalomtudományok szolgáltatta ismereteket is bemenetnek tekintjük a naturalista megfontolásokba, akkor vagy kitágítjuk a „természet” fogalmát, és a hagyományos distinkciókkal szembenelve bele vesszük a szociális világot is,<sup>26</sup> vagy a természet fogalmának eredeti jelentését megtartva elszakadunk a naturalizmus terminus által sugallt értelemről, esetleg elvetve magát a terminust is.

Persze lehet ragaszkodni a természettudományok kitüntettségéhez, és elnevezni a módszertani naturalizmus fogalmának a társadalomtudományok irányába történő kiszélesítését, fenntartva ezzel a két kultúra koncepcióját. Ezt azonban mi éppen a heterogenitást szem előtt tartva nem tartjuk vonzó álláspontnak. Ha ugyanis a természettudományt, sőt már az egyes természettudományokat is a sokféleség jellemzi módszertan, modellek, elköteleződések, érvelési stílusok stb. tekintetében, akkor nem alkotnak olyan egységet, amelybe foglalva szembe lehetne állítani őket a szintén egységesen kezelt társadalomtudományokkal.<sup>27</sup> Az egyes tudományterületeket a „családi hasonlóságokon” túl az intézményes gyakorlat köti össze, és ez alapján nem látunk tartalmas különbséget az általában vett természet- és társadalomtudományok között, amely különbség jellegében eltérne a bármilyen tudományterületek közti különbségektől.

<sup>26</sup> Ez a felfogás jellemző pl. a tudásszociológia erős programjára is, lásd pl. (Bloor 1992, magyarul Bloor 1999a. 428): „A szociológust a tudás – ideértve a tudományos tudást is – kizárólag mint természeti jelenség érdekli.” Egy későbbi fejezetet Bloor „a matematika naturalista megközelítésének” szán, amin azt érti, hogy deskriptív tudományos módszerekkel, vagyis „természeti jelenségeként” kezeli tárgyát. Ebben a megközelítésben természeti az, amit nem normatív, értékutalajdonító módon vizsgálunk, hanem empirikus leírásnak vetünk alá, és oksági nyelvezetben jellemezzük, módszertani folytonosságban a hagyományosan kialakult természettudományokkal.

<sup>27</sup> Érvelésünk ezen a ponton körkörösé válik, hiszen a tudománytanulmányok állításainak tudományos megalapozottságát alátámasztandó felhasználtuk a heterogenitás tézist – a társadalom- és természettudományok közötti elhatárolás feloldásakor –, miközben ezt a tézist a tudománytanulmányok megalapozott állításaként vezettük be. Az érvelés egésze persze ennél komplexebb, így reméljük, hogy ez a körköröség konstruktív jellegű.

Így a naturalizmus fogalmának ez a tágabb, általunk is felvállalt értelmezése azzal a feltevessel él, hogy egy jelenségkör megismerésének legitim birtokosa a rá irányuló tudományterület, függetlenül attól, hogy természeti vagy társadalmi jelenségről van-e szó. Ez a tézis valamiféle folytonosságot feltételez tudományos magyarázatok és naturalista filozófiai magyarázatok között mind a célok, mind a módszerek tekintetében. Ahogy a tudóst, úgy a naturalista filozófust a tudomány juttatja ismereteinek birtokába, így a filozófia és a tudomány közti határvonal elmosódik.

Ennek a nézetnek a bevett interpretációja az, hogy mivel a tudomány empirikus vállalkozás, ezért a naturalista magyarázat is empirikus ismereteken alapul, és az *a priori* elveknek és érveknek nincs helyük a filozófiában (pl. Kitcher 1992). A heterogenitási tézis azonban azt is magában foglalja, hogy az empirikus megismerési mód egyáltalán nem kizárólagos a tudományban. A tudománytörténészek a tudományos gondolkodás és érvelés „stílusairól” beszélnek (Crombie 1994; Hacking 1999), melyek esetén egyáltalán nem egyértelmű és egyöntetű, hogy mi számít empirikus módszernek, és hogyan jelenik meg az empiria a megismerésben. Az elmúlt évtizedekben népszerűvé vált annak vizsgálata, hogy milyen szerepet töltenek be a gondolatkísérletek, vagyis az *a priori* érvelések a tudományokban (Brown 1991; Gendler 2000; Horowitz–Massey 1991). Vannak, akik a gondolatkísérlet jelentőségét abban látják, hogy rácáfol a szigorúan empirista tudományfelfogásra (pl. Brown 1993), míg mások megmentenék az empirizmust azáltal, hogy a gondolatkísérletnek pusztán retorikai (nem pedig tudásbővítő) funkciót tulajdonítanak (Norton 2004). A tudománytanulmányok által követett hagyomány számára kézenfekvő Kuhnra támaszkodni, aki tudásszervező funkciót feltételezve ismeretbővítő, de nem információbővítő jelentőséget szán a gondolatkísérleteknek (Kuhn 1977). Hacsak nem kötelezzük el magunkat egy szigorúan empirista értelmezés mellett, mint pl. Mach tette a múlt században – amelyet a gondolatkísérletekre és azok súlyára irányuló figyelem mértéke kissé valószínűtlenné tesz –, a gondolatkísérletek széles körű jelenléte arra utal, hogy tudomány és filozófia folytonossága kétélű. Nemcsak a filozófia hasonul a tudományhoz annyiban, hogy empirikus megismerési módokból származó tudásra épít (a naturalizmus közkeletű felfogásának megfelelően), hanem a tudomány is hasonul a filozófiához a gondolatkísérletek alkalmazása által.<sup>28</sup>

Így a módszertani heterogenitás elfogadása, és vele az *a priori* tudományos érvelési stílusjegyek elismerése ahhoz vezet, hogy a naturalista tézis kiüresedik. Ha ugyanis a filozófia számos szempontból hasonlít a tudományokra, akkor a filozófia episztémikus függése a tudománytól nem jelent mást, mint hogy a tág értelemben vett megismerési gyakorlatunk bizonyos részei támaszkodnak más részekre, ez pedig akár az egyesítési elméleteknek, akár a redukciós felfogások-

<sup>28</sup> Jó összefoglalóját adja a hipotetikus gondolatmenetek fejlődésének a tudományban Heidelberg–Schiemann 2009.

nak, akár a diszciplináris munkamegosztás nézeteinek közös feltevése. A heterogenitás implikálja, hogy a filozófiának nincs olyan ismertetőjegye, amelyik ne volna jellemző a tudományos gyakorlatok egyikére vagy másikára, és a filozófia intézményes betagozódása a tudomány intézményrendszerébe ezt alátámasztja a kulturális gyakorlat szintjén.

Az állításunk itt nem az a nagyívű tézis, hogy a filozófia tudomány (hiszen éppen a tudomány fogalmának egyértelműségét tesszük kérdésessé), hanem az, hogy a tudományosan alátámasztott tudománykép sokrétűségét tekintve a naturalizmus azon előfeltevése, amelyik episztémikus hierarchiát vagy kapcsolatot, netán munkamegosztást állít tudomány és filozófia között, üresnek tűnik az említett kategóriák fellazulása miatt. Azt sem állítjuk, hogy a filozófus minden tevékenységét érinti ez a gondolatmenet, hiszen csak a (világra vonatkozó) deskriptív állítások megtétele tartozik ide. Ha normatív állításokat tesz a filozófus (természetesen nem a tudomány módszertani normáira gondolunk, hiszen a naturalizmus nem vitatja el ennek jogát a tudósoktól), vagy olyan tevékenységet végez, ahol nincs egyértelmű tudományos szakterület, amely episztémikus prioritást élvezne a filozófiával szemben (mint népi pszichológiai terminusok elemzése),<sup>29</sup> akkor a természettudománytól elkülönül a filozófus tevékenysége – ám pontosan ezeket a tevékenységeket nem szokták naturalizmusként jelölni.

A mindezek alapján értelmezett naturalizmus azt a hitet jelenti, hogy a (természeti, társadalmi, bármilyen) világ megismerésének kitüntetett módja a tudományos megismerés, vagyis egy adott kérdéssel kapcsolatban a legmegbízhatóbb válaszokat a kérdést legitim módon vizsgáló tudományterület nyújtja. Ugyanakkor az egyes tudományterületek episztémikus értelemben autonómak, és az egymás feletti kontrollt a kooperáció szabad lehetősége biztosítja. A kooperáció során alapvető szempont, hogy a területek egymásra támaszkodnak olyan kérdésekben, amelyekben a másik fél rendelkezik kompetenciával. De vajon mit jelent mindez a filozófiára nézve?

## VI. NATURALIZMUS A HETEROGÉN TUDOMÁNY VILÁGÁBAN

A naturalizmus részben a tudomány episztémikus felsőbbrendűségét hirdető filozófusok hatására jött létre, és a társadalmi intézményként megerősödő tudomány tudástermelő autoritásigénye ma már állásfoglalásra készíti a legtöbb

<sup>29</sup> A népi pszichológia (folk psychology) a filozófiai hagyományban részben pont az empirikus adatgyűjtésen alapuló pszichológia autoritását kikerülendő jelent meg, módszertanában eredetileg kiemelve az introspekciót. Kicsit bonyolultabban, de az ontológiai naturalizálásnál ismertetett probléma itt is megjelenik. A mentális meghatározása ugyanakkor problematikus. Ennek a naturalizmussal konzisztens kísérletéhez lásd Demeter 2009, illetve ennek kritikáját (Tózsér 2010), amely „üresnek” tartja a mentális valóságot kezelését a halmazba tartozási kritériumok megadása nélkül.

filozófiai hagyományt. A filozófia reakciója sokféle lehet a tudományok felé fordulástól és egyre intenzívebb közös munkától kezdve (ahogy erre a biológia-filozófia, a kognitív tudományok és a fizika egyes részdiszciplínái kitűnő példát adnak) a tudománytól való elforduláson át a filozófiai hagyományba való befordulásig (újrateremtve az iskolafilozófiát). Mi a tudomány heterogenitásának jelentőségét e két utóbbi hozzáállással kapcsolatban nem vizsgáltuk, és az elsővel kapcsolatban is inkább negatív konklúziókhöz jutottunk. Amellett érveltünk, hogy az ontológiai naturalizmus alapfeltevései problematikusak, a módszertani naturalizmus pedig kiüresedik a mai tudománytanulmányok fényében. De kínál-e ennek a heterogenitásnak a felismerése pozitív tanulságokat a tudomány-nyal szorosán együttműködni kívánó filozófia számára (ha a fentiek fényében érdemes más elnevezést találni a naturalizmus helyett)? Hitünk szerint igen, és a következőkben nagyon röviden megpróbálunk néhány példát adni a heterogén tudomány világában jelentkező releváns filozófiai problémákra.

E kérdések egy része az episztémikusan kitüntetett tudomány történetileg kialakult szerveződéséből fakad. Ha azt gondoljuk, hogy a világról minden eddiginél többet tudunk, és azt jobban (megbízhatóbban, pontosabban, árnyaltabban) tudjuk, akkor fontos kérdés, hogy milyen módon jutottunk és jutunk ehhez a tudáshoz. Ennek vizsgálatakor az egyik legérdekesebb probléma magának a tudásnak a kérdése. A filozófiai hagyomány általában a tudást individuális megismerő ágensek kapcsán vizsgálja. Ha azonban megnézzük a tudományos tudást létrehozó struktúrákat, egyre inkább kollektív folyamatnak tűnik a tudástermelés.<sup>30</sup> Az episztémikus függés megjelenik egyrészt a primer tudástermelésben (Hardwig 1985), de megjelenik a kollektívák által létrehozott tudás átadásában is.<sup>31</sup> Míg tudásnak ma is elsősorban a személyek tudását tartjuk, a társadalmunkban kitüntetett tudást közösségek termelik és veszik át egymástól – egy olyan folyamatban, ahol a szakértői közösségek – a hatékonyabb tudástermelés igé-

<sup>30</sup> Ez a felismerés ha lassan is, de a tudományt az individuális kognitív ágensek mentén értelmező naturalista filozófiában is megjelenik. Még Ronald Giere is, aki korábbi munkáiban az erősödő szociologizálás alternatívájának tartotta a kognitív megközelítést (Giere 1988), ma már efelé hajlik, hogy „[a]z utóbbi években a kognitív tudományon belül többen elismerték, hogy a megismerésnek irreducibilis externális és társadalmi komponense van” (Giere 2008, 262). Egyben maguk a kognitív modellek is egyre nyilvánvalóbbá teszik ezt az elköteleződést, mind a tudomány elemzésénél (Carruthers–Stich–Siegal 2002; Giere–Moffatt 2003; Gorman 2005; Hutchins 1995; Magnus 2007), mind a kognitív ontogenetikai és filogenetikai fejlődésének vizsgálatakor (Csibra–Gergely 2009; Donald 2001; Sperber 2001; Tomasello 1999), hogy csak néhány példát említsünk.

<sup>31</sup> A tudományos tudás kialakulásával foglalkozók számára ennek az episztemológiai kérdésnek a fontossága korán nyilvánvalóvá vált, ahogy a sokáig elfeledett Ludwik Fleck már a harmincas években megjegyezte: „A populáris tudomány minden ember tudásának jelentős részét adja, még a legegztakabb szakterületek szakértői is számos fogalmat, számos összehasonlítási szempontot és általános nézőpontot köszönhetnek ennek, emiatt a populáris tudomány minden megismerésünk során figyelembe veendő episztemológiai probléma” (Fleck 1980: 148).

nyeinek megfelelően – nincsenek tökéletes ismeretében más közösségek nyelvhasználati (és szemantikai) elköteleződéseinek.<sup>32</sup>

Számos filozófiailag releváns probléma következik ebből. Ilyen a tudás értelmezésének kérdése, amely kapcsán még mindig dominánsnak tűnik a logikai pozitivisták hagyomány erőteljes szemantikai (és szintaktikai) elköteleződése, amely háttérbe szorította a pragmatikai kérdéseket és azt a problémát, hogy más a hozzáférése a jelentéshez a filozófusnak és a szaktudósoknak. De ilyen a tapasztalat és következtetés jelentőségének újragondolása is, amelyek ugyan hagyományosan a megbízható tudás garanciáinak számítottak, ám jellemzésük során eltekintettek a közlés (testimónium) mára már nyilvánvalóvá vált alapvető jelentőségétől. A tudás fogalmának kollektivisták újraértelmezése mellett előtérbe helyezi mind a „nyelvjátékok”, „lexikonok”, vagy egyéb, konstitutívnak tekintett struktúrák szerepét, mind a cselekvési formákkal átadható hallgatolagos komponensek jelentőségét. Egy másik, de kapcsolódó problémakör a heterogenitás kapcsán annak a disszenzusnak a filozófiai jelentősége, melynek jelenléte a tudományban már a laikus számára is egyre inkább nyilvánvaló. A szociológiai elméletek általában a konszenzust tekintik a tudás megbízhatóságának mércéjeként, azonban kétségtelen, hogy a konszenzus nem garantál tudást a koherenciaelméleteken túlmutató értelemben, és feltételezhető, hogy ez a diszkrépancia hozzájárult a két hagyomány eltávolodásához.

A heterogenitást elismerő naturalizmus ugyanakkor nemcsak az ismeretelmélet, hanem az ontológia számára is tanulságokkal szolgál. Éppen az a tény, hogy a tudomány nem kínál egységes, vagy akár koherens ontológiai válaszokat az erre építendő filozófiai világképek vagy érvelések számára, filozófiai szempontból érdekes és vizsgálatra szorul. Ugyanakkor a fundamentálonológia tagadása – beleértve ebbe a tentatív és történetileg kontingens, ám aktuálisan mégis átfogó ontológia lehetőségének tagadását is – nem vezet ontológiai anarchizmushoz annyiban, hogy tudományos kutatással feltérképezhető a tudomány ontológiai elköteleződéseit irányító mechanizmusok, és ezek fontos információval szolgálhatnak a relativista vagy historicista ontológiák számára is.

Természetesen tudjuk, hogy ezek a javaslatok nem forradalmian újak. A társas episztemológiák elkezdtek felmérni a tudományos tudás közösségi és individuális összjátékából származó hibridjének filozófiai kérdéseit, az empirikus filozófia elkezdte elemezni a filozófiai megismerés sokáig irreflexíven kezelt módszereit, a szaktudományok filozófusai konstruktívan dolgoznak együtt tudósokkal, számos filozófus inspiratívan összegzi és popularizálja a tudományos eredményeket, fontos összekötő kapcsot létesítve a különböző háttértudású kö-

<sup>32</sup> A filozófia számára így a naturalizálás kapcsán nagyon közvetlenül merül fel a szakértelem kérdése: milyen módon használja fel a tudomány tudását? Mennyire legitim az, amit mond? Ennek a kérdésnek sok szempontját érintette a magyar kontextusban a „genetizmus” kapcsán kibontakozó vita (Boros–Guttman 2004a és 2004b, Nemes–Molnár–Kakuk 2004).

zösségek között. A filozófia virágzik, és számos ponton tud gyümölcsözően kapcsolódni a tudományhoz.

Ám a fenti területeken a kutatás még épphogy csak elkezdődött. Ahogy a viszonylag fiatal társadalomtudományok egyelőre távolabb állnak a felségterületekre tartozó jelenségek köreinek átfogóbb magyarázataitól, mint a régebben intézményesült és nagyobb támogatásnak örvendő természettudományok, úgy a tudomány kutatására szakosodott diszciplína is csak relatíve kevés tankönyvszintűen összefoglalható ismeretet állított elő néhány évtizedes fejlődése során. Ez a diszciplináris nyitottság és elméleti rugalmasság azonban annak jele, hogy a terület számára gyümölcsöző lehet a filozófiai reflexió, a magától értetődőnek tekintett elköteleződések felülvizsgálata, az alapvető fogalmak és modellek elemzése – sokszor akár az esettanulmányok burjánzó sokaságának rovására is.<sup>33</sup> Így a területek közti kooperáció kétirányú lehet, hiszen a viszonylag friss tudományterület éppúgy rászorul a tárgyát hagyományosan vizsgáló filozófia támogatására, mint ahogy ez a filozófia rászorul a számára releváns empirikus ismeretek növekvő körének tekintetbevételére. Ez pedig nem meglepő: a tudomány és filozófia közti határvonal elmosásával, a folytonosság feltételezésével valószínűtlenné tettük az olyan hierarchikus viszonyokat, mint amilyeneket egy szigorú naturalista program elképzelt a tudományoknak alárendelt filozófia esetén.

## IRODALOM

- Almeder, Robert 1993. On Naturalizing Epistemology. In James H. Fetzer (szerk.) *Foundations of Philosophy of Science*. New York, Paragon House. 451–473.
- Asquith, P. L. – I. Hacking (szerk.) *Proceedings of the 1978 Meeting of the Philosophy of Science Association*. 197–214. East Lansing, The Philosophy of Science Association.
- Barnes, Barry – David Bloor – John Henry 2002. *A tudományos tudás szociológiai elemzése*. Ford. Faragó Péter – Tanács János. Budapest, Osiris.
- Bechtel, William 2006. *Discovering Cell Mechanisms: The Creation of Modern Cell Biology*. Cambridge, Cambridge University Press.
- Bechtel, William 2008. *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. London, Routledge.
- Bloor, David 1992. *Knowledge and Social Imagery*. London, Routledge and Kegan Paul.
- Bloor, David 1999a. A tudásszociológia erős programja. Ford. Farkas Katalin. In Forrai–Szegei 1999. 427–445.
- Bloor, David 1999b. Anti-Latour. *Studies in History and Philosophy of Science*. Part A 30. 81–112.

<sup>33</sup> Az esettanulmányok túlzott tobzódása a tudománytanulmányok jelenlegi hibájaként is felrható. Míg a kezdeti erőfeszítéseket az elméleti keretek keresése jellemezte, gyakran nyitottan filozófiai megfontolásokra, addig a későbbiekben a hangsúly egyre inkább eltolódott a lokális esetek empirikus elemzése felé. Ennek részben az is oka lehet, hogy a súlyosabb belső viták többnyire nem konstruktívan, az álláspontok konvergenciájával végződtek (pl. Collins–Yearly 1992; Bloor 1999b), és a szakma az elméletalkotástól a deskriptív munka felé fordult.

- Boros János – Guttman András 2004a. Genetizmus: gének és társadalom. *Magyar Tudomány*. 49/6. 752–756.
- Boros János – Guttman András 2004b. Gének, társadalom és az értelmezés tudománya. *Magyar Tudomány*. 49/9. 1029–1032.
- Brigandt, I. – A. Love, 2008. Reductionism in Biology. In Edward N. Zalta (szerk.) *The Stanford Encyclopedia of Philosophy*. URL = <http://plato.stanford.edu/entries/reduction-biology>.
- Brown, James Robert 1991. *Laboratory of the Mind: Thought Experiments in the Natural Sciences*. London, Routledge.
- Brown, James Robert 1993. Why Empiricism Won't Work. *Proceedings of the Philosophy of Science Association*. 2. 271–279.
- Carruthers, Peter – Stephen P. Stich – Michael Siegal 2002. *The Cognitive Basis of Science*. Cambridge, Cambridge University Press.
- Cartwright, Nancy – Jordi Cat – Lola Fleck – Thomas E. Uebel 1996. *Otto Neurath: Philosophy between Science and Politics*. Cambridge, Cambridge University Press.
- Chadarevian, Soraya de – Nick Hopwood, 2004. *Models: The Third Dimension of Science*. Stanford, Stanford University Press.
- Chang, Hasok megj. alatt. *Is Water H<sub>2</sub>O? Evidence, Pluralism and Realism*. Dordrecht, Springer.
- Clark, Thomas W. 2007 *Encountering Naturalism: A Worldview and Its Uses*. Somerville, Center for Naturalism.
- Collins, Harry – Robert Evans 2002. The Third Wave of Science Studies: Studies of Expertise and Experience. *Social Studies of Science*. 32. 235–296.
- Collins, Harry – Robert Evans 2007. *Rethinking Expertise*. Chicago, The University of Chicago Press.
- Collins, Harry – Steven Yearly 1992. Epistemological Chicken. In A. Pickering (szerk.) *Science as Practice and Culture*. Chicago, Chicago University Press. 301–326.
- Crane, Tim – D. H. Mellor 1990. There is No Question of Physicalism. *Mind*. New Series. 99/394. 185–206.
- Craver, Carl F. 2007. *Explaining the Brain*. Oxford, Oxford University Press.
- Crombie, Alistair C. 1994. *Styles of Scientific Thinking in the European Tradition*. London, Duckworth.
- Csibra Gergely – Gergely György 2009. Natural Pedagogy. *Trends in Cognitive Sciences*. 13. 148–153.
- Demeter Tamás 2009. Two Kinds of Mental Realism *Journal of General Philosophy of Science*, 40. 59–71.
- Donald, Merlin 2001. *Az emberi gondolkodás eredete*. Ford. Kárpáti Eszter. Budapest, Osiris.
- Duhem, Pierre Maurice Marie 1954. *The Aim and Structure of Physical Theory*. Princeton, Princeton University Press.
- Dunbar, Kevin 2002. Understanding the Role of Cognition in Science: The Science as Category Framework. In P. Carruthers. – S. Stich – M. Siegal (szerk.) *The Cognitive Basis of Science*. Cambridge, Cambridge University Press. 154–170.
- Dupré, John 1993. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, Mass., Harvard University Press.
- Falkenburg, Brigitta 1996. The Analysis of Particle Tracks: A Case for Trust in the Unity of Physics. *Studies in History and Philosophy of Modern Physics*. 27. 337–371.
- Farrell, Robert P. 2003. *Feyerabend and Scientific Values: Tightrope-Walking Rationality*. Dordrecht, Kluwer.
- Feldman, R. 2006. Naturalized Epistemology. In Edward N. Zalta (szerk.) *The Stanford Encyclopedia of Philosophy* (Fall 2006 Edition). *The Stanford Encyclopedia of Philosophy* URL = <http://plato.stanford.edu/archives/fall2006/entries/epistemology-naturalized/>

- Fleck, Ludwik 1980. *Entstehung und Entwicklung einer wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkollektiv*. Ford. Lothar Schäfer – Thomas Schnelle. Frankfurt am Main, Suhrkamp.
- Fodor, Jerry A. 1974. Special Sciences, or the Disunity of Science as a Working Hypothesis. *Synthese*. 28. 77–115.
- Forrai Gábor – Szegedi Péter (szerk.) 1999. *Tudományfilozófia szöveggyűjtemény*. Budapest, Áron Kiadó.
- Frankel, H. 1978. The Non-Kuhnian Nature of the Recent Revolution in the Earth Sciences. P. L. Asquith – I. Hacking (szerk.) *Proceedings of the 1978 Meeting of the Philosophy of Science Association*. East Lansing, The Philosophy of Science Association. 197–214.
- Galison, Peter – David J. Stump 1996. *The Disunity of Science: Boundaries, Contexts, and Power*. Stanford, Stanford University Press.
- Galison, Peter 1987. *How Experiments End*. Chicago, University of Chicago Press.
- Galison, Peter 1997. *Image and Logic: a Material Culture of Microphysics*. Chicago, University of Chicago Press.
- Gendler-Szabo, Tamar 2000. *Thought Experiment: On the Powers and Limits of Imaginary Cases*. New York, Garland Press.
- Giere, Ronald N. 1985. Philosophy of Science Naturalized. *Philosophy of Science*, 52. 331–356.
- Giere, Ronald N. 2008. Cognitive Studies of Science and Technology. In Edward J. Hackett – Olga Amsterdamska – Michael Lynch – Judy Wajcman (szerk.) *The Handbook of Science and Technology Studies*. Cambridge, Mass., The MIT Press. 259–278.
- Giere, Ronald N. – Barton Moffatt 2003. Distributed Cognition: Where the Cognitive and the Social Merge. *Social Studies of Science*. 33. 301–310.
- Giere, Ronald. 1988. *Explaining Science: A Cognitive Approach*. Chicago, University of Chicago Press.
- Gieryn, Thomas F. 2006. Határmunkálatok és a tudomány elhatárolása a nem-tudománytól: feszültségek és érdekek a tudósok szakmai ideológiáiban. Ford. Lukács Anikó. *Replika*. 54–55. 173–194.
- Golinski, Jan. 1998. *Making Natural Knowledge: Constructivism and the History of Science, Cambridge History of Science*. Cambridge, Cambridge University Press.
- Gorman, Michael E. 2005. Levels of Expertise and Trading Zones: Combining Cognitive and Social Approaches to Technology Studies. In Michael E. Gorman – Ryan Tweney – David Gooding – Alexandra Kincannon (szerk.) *Scientific and Technological Thinking*. Mahwah, NJ, Lawrence Erlbaum. 287–302.
- Gross, Paul – Norman Levitt 1994. *Higher Superstition: The Academic Left and Its Quarrels with Science*. New Haven, Yale University Press.
- Hacking, Ian 1983. *Representing and Intervening*. New York, Cambridge University Press.
- Hacking, Ian 1999. A tudományos gondolkodás stílusa. Ford. Demeter Tamás. In Forrai-Szegedi 1999. 349–365.
- Hardwig, John 1985. Epistemic Dependence. *The Journal of Philosophy*. 82. 335–349.
- Heidelberger, Michael – Gregor Schiemann 2009. *The Significance of the Hypothetical in Natural Science*. Berlin – New York, Walter de Gruyter.
- Henke, Christopher R. – Thomas F. Gieryn 2008. Sites of Scientific Practice: The Enduring Importance of Place. In Edward J. Hackett – Olga Amsterdamska – Michael Lynch – Judy Wajcman (szerk.) *The Handbook of Science and Technology Studies*. Cambridge, Mass., The MIT Press. 354–376.
- Hersh, Reuben 1979. Some Proposals for Reviving the Philosophy of Mathematics. *Advances in Mathematics*. 31. 31–50.
- Horowitz, Tamara – Gerald J. Massey 1991. *Thought Experiments in Science and Philosophy, CPS Publications in Philosophy of Science*. Savage, Rowman & Littlefield.



- Hutchins, Edwin 1995. *Cognition in the Wild*. Cambridge, MA, MIT Press.
- Kitcher, Philip 1992. The Naturalists Return. *Philosophical Review*. 101. 53–114.
- Kitcher, Philip 1998. 1953 and All That: A Tale of Two Sciences. In M. Curd – J. A. Cover (szerk.) *Philosophy of Science: Central Issues*. New York, W.W. Norton & Co. 971–1003.
- Knorr-Cetina, Karin 1999. *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, Mass., Harvard University Press.
- Kornblith, Hilary 1985. What Is Naturalistic Epistemology. In Kornblith, Hilary (szerk.) *Naturalizing Epistemology*. Cambridge, Mass., MIT Press. 1–13.
- Kuhn, Thomas S. 1977. A Function for Thought Experiments. In *The Essential Tension*. Chicago, University of Chicago Press. 240–265.
- Kuhn, Thomas S. 1990. The Road Since Structure. *PSA*. 2. 3–13.
- Kuhn, Thomas S. 1993. Afterwords. In Paul Horwich (szerk.) *World Changes: Thomas Kuhn and the Nature of Science*. Cambridge, MA, MIT Press. 311–341.
- Kusch, Martin 1995. *Psychologism. A Case Study in the Sociology of Philosophical Knowledge*. London, Routledge.
- Kutrovácz Gábor – Láng Benedek – Zemplén Gábor 2010. Bevezető. In uőök (szerk.) *Határmunkálatok a tudományban*. Budapest, L'Harmattan 7–20.
- Kutrovácz Gábor 2010a. Szakértő laikusok, laikus szakértők. In Csörgő Zoltán – Szabados Levente (szerk.) *Szubjektív tudás – objektív tudomány. A XVI. Magyar Kognitív Tudományok Konferencia publikációinak gyűjteménye*. Budapest. 194–206.
- Kutrovácz Gábor 2010b. Hogyan értékelhetünk szakértői véleményeket? In Kutrovácz Gábor – Láng Benedek – Zemplén Gábor (szerk.) *Határmunkálatok a tudományban*. Budapest, L'Harmattan. 148–165.
- Laki János 2010. Hitek, módszerek, megbízhatóság. Mit demarkál a kritérium? In Kutrovácz Gábor – Láng Benedek – Zemplén Gábor (szerk.) *Határmunkálatok a tudományban*. Budapest, L'Harmattan. 166–184.
- Latour, Bruno – Woolgar, Steve 1986. *Laboratory Life: the Construction of Scientific Facts*. Princeton, Princeton University Press.
- Latour, Bruno 2004. The Last Critique. *Harper's Magazine*. April. 15–20.
- Laudan, Larry. 1983. The Demise of the Demarcation Problem. In R. Cohen – L. Laudan (szerk.) *Physics, Philosophy and Psychoanalysis*. Dordrecht, Reidel. 111–128.
- Laudan, Larry. 1999. A tudomány áltudománya. Ford. Farkas Katalin. In Forrai-Szegedi 1999. 477–495.
- Laudan, Rachel. 1980. The Recent Revolution in Geology and Kuhn's Theory of Scientific Change. In Garry Gutting (szerk.) *Paradigms and Revolutions. An Interdisciplinary Approach to Kuhn*. South Bend, Notre Dame University Press. 284–296.
- LeGrand, H. E. 1988. *Drifting Continents and Shifting Theories: the Modern Revolution in Geology and Scientific Change*. Cambridge, Cambridge University Press.
- Lloyd, Elisabeth A. 1997. Feyerabend, Mill, and Pluralism. *Philosophy of Science*. 64. Supplement. *Proceedings of the 1996 Biennial Meetings of the Philosophy of Science Association. Part II: Symposia Papers*. December. 396–407.
- Lynch, Michael – Steve Woolgar (szerk.) 1990. *Representation in Scientific Practice*. Cambridge, Mass., MIT Press.
- Lynch, Michael. 2006. Expert Metascientists. *Social Studies of Science*. 36/6. 867–868.
- Magnus, P. D. 2007. Distributed Cognition and the Task of Science. *Social Studies of Science*. 37/2. 297–310.
- Morris, Charles 1960. On the History of the International Encyclopedia of Unified Science. *Synthese*. 12. 517–521.
- Morrison, Margaret 2000. *Unifying Scientific Theories*. Cambridge, Cambridge University Press.

- Nemes László – Molnár Péter – Kakuk Péter 2004. Több dolgok földön és égen. *Magyar Tudomány*. 9. 1033–1038.
- Nisbett, Richard E. 2003. *The Geography of Thought: How Asians and Westerners Think Differently – and Why*. New York, The Free Press.
- Norton, John. 2004. Why Thought Experiments Do Not Transcend Empiricism. In C. Hitchcock (szerk.) *Contemporary Debates in the Philosophy of Science*. Oxford, Blackwell. 44–66.
- Oppenheim, Paul – Hilary Putnam 1958. Unity of Science as a Working Hypothesis. *Minnesota Studies in the Philosophy of Science*. 2. 3–36.
- Papineau, David. Naturalism. In Edward N. Zalta (szerk.) *The Stanford Encyclopedia of Philosophy*. URL = <http://plato.stanford.edu/entries/reduction-biology>.
- Quine, W. van Orman 1999. Naturalizált ismeretelmélet. Ford. Farkas Katalin. In Forrai–Szegegyi 1999. 369–382.
- Rheinberger, Hans-Jörg 1997. *Toward a History of Epistemic Things: Synthetizing Proteins in the Test Tube*. Stanford, Stanford University Press.
- Rheinberger, Hans-Jörg 2006. *Epistemologie des Konkreten: Studien zur Geschichte der modernen Biologie*. Frankfurt am Main, Suhrkamp.
- Rheinberger, Hans-Jörg 2010. *Historical Epistemology*. Palo Alto, Stanford University Press.
- Shapin, Steven 2006. Hogyan legyünk tudományellenesek? *Replika*. 3. 23–36. 157–171.
- Sokal, Alan – Jean Bricmont 2000. *Intellektuális impostorok*. Ford. Kutrovácz Gábor. Budapest, Typotex.
- Solomon, Miriam 1992. Scientific Rationality and Human Reasoning. *Philosophy of Science*. 59/3. 439–455.
- Solomon, Miriam 2001. *Social Empiricism*. Cambridge, Mass., The MIT Press.
- Solomon, Miriam 2006a. *Groupthink* versus *The Wisdom of Crowds*: The Social Epistemology of Deliberation and Dissent. *The Southern Journal of Philosophy*. 44. 28–42.
- Solomon, Miriam 2006b. Norms of Epistemic Diversity. *Episteme*. 3. 23–36.
- Sperber, Dan 2001. *A kultúra magyarázata*. Ford. Pléh Csaba. Budapest, Osiris.
- Star, S. L. – J. R. Griesemer 1989. Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–1939. *Social Studies of Science*. 19/4. 387–420.
- Stewart, J. 1990. *Drifting Continents and Colliding Paradigms*. Bloomington, Indiana University Press.
- Sunstein, Cass R. 2003. *Why Societies Need Dissent*. Cambridge, Mass., Harvard University Press.
- Thagard, Paul – Gregory Nowak 1990. The Conceptual Structure of the Geological Revolution. In J. Shrager – P. Langley (szerk.) *Computational Models of Scientific Discovery and Theory Formation*. San Mateo, Morgan Kaufman. 27–72.
- Tomasello, Michael 1999. *The Cultural Origins of Human Cognition*. Cambridge, MA, Harvard University Press.
- Tőzsér János 2009. Mental Realism Reloaded. *Journal for General Philosophy of Science*. 40/2. 337–340.
- Vermeir, Koen 2006. Újra a demarkációs problémáról. Ford. Koroncay Dávid – Zelei Borbála. *Replika*. 54–55. 135–156.
- Zemplén Á. Gábor 2010. Kognitív munkamegosztás és a tudás termelése. In Csörgő Zoltán – Szabados Levente (szerk.) *Szubjektív tudás – objektív tudomány. A XVI. Magyar Kognitív Tudományok Konferencia publikációinak gyűjteménye*. Budapest. 182–193.

Számunk *Fórum* rovatának fókuszpontjában Laki János *A tudomány természete* (Budapest, Gondolat, 2006) című kötete áll. Azok az oldalszámok, amelyek a szerző által adott összefoglalásban, az ezt követő recenziókban (Pléh Csaba, Zemplén Gábor, Forrai Gábor és Schwendtner Tibor munkái), valamint Laki János válaszában a szerző nevének megadása nélkül szerepelnek, kivétel nélkül erre a kötetre vonatkoznak. A recenziók és a válasz szövegét terjedelmi okokból rövidített formában közöljük, reményeink szerint azonban a vitakör teljes anyaga hamarosan felkerül a *Szemle* honlapjára. A *Fórum* összeállításáért köszönettel tartozunk Demeter Tamásnak.



## Tapasztalat, nyelv, társadalom, tudományfejlődés – Kuhn még egyszer\*

Ahogy a filozófiatörténet számos más alakja, Thomas Kuhn is szeretett úgy tekinteni önmagára, mint félreértett gondolkodóra. Bírálóival gyakran volt türelmetlen, s élete különböző szakaszaiban következetesen hangsúlyozta, írásait félreértik, olyan nézeteket és szándékokat tulajdonítanak neki, melyek sohasem voltak övéi. Egyszer egyenesen azt írta, úgy érzi, mintha két Kuhn lenne: egy, aki az ő műveit írta, s egy másik, aki azonos néven, azonos című és betűjű, de egészen más értelmű munkákat publikált. A tudományháborúktól sújtott utókor vagy úgy tekint rá, mint az egyik első számú ellenségre, vagy úgy, mint a tudomány indokolhatatlan kognitív autoritása elleni felszabadító harc hőisére. Mindkét tábor *negatív* eredményeket tekint Kuhn fő, sőt egyetlen érdemének: a tudomány naív empirista, elementarista, fundacionalista ismeretelméleti alapjainak diszkreditálását, s ezzel összefüggésben a jelentés igazságfeltétel-elméletének, a behaviorista elme-felfogásnak, a tudomány tudásszociológiai vizsgálat alóli fölmentésének vagy a prezentista tudománytörténet-írásnak az ellehetetlenítését. Az én könyvem főszereplője az a Kuhn, akinek újszerű elgondolása volt a tudomány racionalitásáról, társadalmi jellegéről, az elméletváltásról, a tudomány történeti fejlődéséről, a realizmusról és a relativizmusról. Céloom annak a nem elég világosan és részletesen megírt tudományfelfogásnak a rekonstruálása volt, amelyet a „bevett nézetet” bíráló Kuhnról alkotott kép sokáig elfedett.

Ahogy célkitűzésem, úgy választott módszerem sem követte a Kuhn-mono-gráfiák szokásos eljárását: a „*jóindulatú anakronizmus*” jegyében lemondtam a radikális tézisek utólagos kozmetikázásának, a legszélsőségesebb tételek átértelmezéseinek vagy súlytalanítási kísérleteinek szorgos leltárba vételéről. Úgy tekintettem a hozzáférhető szövegekre, mint egységes korpuszra, melyből egy részleteiben némiképp változó, de alapvető intencióját tekintve változatlan, s mindaddig kevés figyelemre méltatott, *pozitív* elképzelés bontakozik ki a tudo-

\* Bevezető *A tudomány természete. Thomas Kuhn és a tudományfilozófia történeti fordulata*. Budapest, Gondolat. 2006 című könyvem megvitatásához. Köszönettel tartozom az Országos Tudományos Kutatási Alapnak kutatásaim támogatásáért (OTKA K 72598).

mány természetéről. Ebben a fényben az utólagos módosítások a kritikák által jelzett félreértések tisztázásai, s nem elhamarkodott vélemények nyílt vagy hallgatólagos visszavonásának dokumentumai.

Kiindulópontom az volt, hogy Kuhn pozitív tudományfelfogása két okból vált jóformán „láthatatlanná”.

1. A fő ok a kortársak (*nem* a neopozitivisták, inkább a popperianusok) által kialakított „*standard interpretáció*”, mely szerint Kuhn irracionális elméletváltásokkal szabdalt változást, de nem fejlődési folyamatként írta le a tudomány történetét. Az összemérhetetlenség egyúttal összehasonlíthatatlanságot és értelmezhetetlenséget is jelent, s mivel a paradigmák maguk konstruálják empirikus bázisukat, sem az igazolás/cáfolás, sem a realizmus bármily gyöngé változata, sem a teoretikus terminusok jelentésének megfigyelések általi korlátozása nem jöhet szóba. A logikai következtetéseket és empirikus ellenőrzést retorikai fogások és pszichológiai manipulációk helyettesítik, így a tudomány szélsőségesen relativista, irracionális vállalkozás.
2. Az értelmezések félresiklásának másik fontos okaként a tudományról a 20. század első felében kialakított „*bevett nézet*” helyzetének téves felmérését jelöltem meg. Nagyjából a 80-as évek végéig úgy látszott, hogy a logikai pozitívizmus nyomán kialakult, egyértelmű dogmákban és dichotómiákban megfogalmazott tudományfelfogást az 50-es évek végéig úgyszólván teljes egyetértés övezte, így Kuhn nézetei hirtelen alakváltásként (Gestalt-switch) tűntek föl a tudományfilozófiában. Ezért került a figyelem középpontjába a pozitív tudományfelfogás helyett az a negatív kérdés, hogyan cáfolta Kuhn az elődöket és kortársakat.

A könyv nagyobb részben a „standard interpretáció” tarthatatlanságáról szól, de a megfelelő helyeken vázoltam a „bevett nézet” ismeretelméleti, szemantikai, pszichológiai és tudásszociológiai dogmáinak a század közepén kezdődött lebomlási folyamatát, jelezve, hogy Kuhn nézetei e folyamatba illeszkednek. Természetesen nem azt akartam megmutatni, hogy nincsenek alapvető különbségek a Kuhn előtti és utáni tudományfilozófia között, csupán azt, hogy e különbségek nem egy csapásra jöttek létre, hanem apró módosulások sorozataként, s ha ezekről nem veszünk tudomást, csak két ideáltípust hasonlítunk össze, s nem látjuk a tényleges történeti átalakulás folyamatát.

Kuhn pozitív tudományfelfogásának kifejtését két nagy részre, s ezek alfejezeteire tagoltam:

- I. A „Tapasztalat” és a „Nyelv” fejezetekből álló első rész Fodor, illetve Putnam arra irányuló kísérleteit vizsgálja, hogy helyreállítsák az empirikus-teoretikus dichotómia valamilyen változatát, s ezáltal megmentsék a tudomány racionalitását, cáfolják az inkommenzurabilitást, elhárítsák a relativizmust, és értelmezhetővé tegyék a tudomány fejlődését.

II. A „Társadalom” főcímet viselő második részben tárgyaltam a paradigma-váltások mikrofolyamatait, valamint a tudomány egészének fejlődését jellemző makrofolyamatot, megmutatván, hogy Kuhn szociológiai koncepciója a semleges tapasztalat feltételezése nélkül is megőrzi a tudomány racionalitását és progresszivitását, s elkerüli a szélsőséges relativizmus csapdáját.

## I. AZ EMPIRIKUS-TEORETIKUS DICHOTÓMIA

A standard interpretáció szerint, Kuhn katasztrofális következményekkel járó állítása az lenne, hogy a tudományos *elméletek* generálják önmaguk empirikus bázisát, illetve tüntetik el a falszifikáló adatokat. Ezzel a fantommal száll szembe Fodor moduláris architektonikáról szóló teóriája, melynek központi tézise, hogy a biológiai evolúció során az elmében a legegyszerűbb, automatikus következtetésekre korlátozódó, a magasabb kognitív szintek számára áthatolhatatlan, ezért gyors és buta, de veridikus észleleteket szolgáltatató neurális modulok jöttek létre. Ezek mindenki számára azonos empirikus anyagot gyűjtenek, melyet a magasabb kognitív szintekhez tartozó funkciók utólag finomítanak, pontosítanak. Viták esetén vissza lehet térni a modulok objektív adataihoz, s ez racionálisan magyarázhatóvá teszi az elméletek elfogadását, korrigálását vagy elvetését.

A modulokból származó tapasztalatok objektivitását Fodor két érveléssel támasztja alá:

1. Az észlelés lehetővé teszi a túlélést, s ez bizonyítja, hogy az veridikusan reprezentálja a környezetet. Könyvem tapasztalatról szóló részének egyik fontos érve, hogy a túlélés nem a környezet egyedül helyes leképezését, csak a perceptuális reprezentáció *empirikus adekvátóságát* bizonyítja. A kémiai vagy mágneses jelekre, ultrahangra vagy elektromos ingerekre érzékeny állatok egyaránt képesek a túlélésre, azaz adekvát módon észlelik környezetüket, de erősen különböző észleleti világokban, ökológiai fülkékben élnek. A túlélés nem a reprezentációk *igaz* voltát bizonyítja, csak azt, hogy az adott faj *sikeresen* választja ki és szervezi egységgé a környezetnek életmódja fenntartásához fontos jellemzőit. A kognitív sikerességet empirikusan adekvát, de perceptuálisan összemérhetetlen reprezentációk is biztosíthatják.
2. A tapasztalatok és hitek elszigeteltségét demonstráló másik fodori érvelésből sem következik az észleletek uniformitása. A kognitív áthatolhatatlanság csak azzal a pótlólagos föltevessel eredményez objektivitást, hogy a kauzálisan meghatározott ingerek és az észlelési apparátus fiziológiája elégséges a tapasztalat természetes egységéhez. Mondandóm röviden az volt, hogy Fodor érve lehet hatásos a tapasztalat *elmélettel* való átitatottságát képviselő tudományfilozófusokkal szemben, de nem találja el Kuhnt, aki nem az em-

pirikus anyag utólagos teoretikus értelmezéséről, hanem magának az észlelésnek begyakorolt sémák általi vezérléséről beszél. Ezt megmutatandó, az észlelőapparátus és a magasabb kognitív képességek együttműködésének két típusát különböztettem meg. A *helmholzi* a hiányos információkat és rövid észlelési időket korábbi tapasztalatokon alapuló, nem tudatos következtetések alkalmazásával *egészíti ki*, míg a *jamesi* észlelésmodell fő jellemzője a túlsúfolt és kaotikus ingeranyag *szelektálása-rendezése*. Úgy látom, Fodor helmholzi jellegűnek tekinti az észlelést, s amellet érvel, hogy ha a proximális ingerekből a modulok elméletek közbejötté nélkül következtetnek a dolgok disztális elrendeződésére, akkor az észlelet veridikus, ha viszont elmélet jön közbe, akkor *illúzió*. Kuhn ellenben a jamesi szelektáló-rendező modellt alkalmazza, melyben az empirikus különbségek nem explicit *elmélet*, hanem a szelektálást-rendezést irányító *beállítódás* következményei. Az ingerek földolgozását *mentális modulok* végzik, melyek „modulok”, amennyiben az idegrendszer részei, de „mentálisak”, amennyiben működésüket mintakövetés és gyakorlás általi programozás határozza meg. Az összemérhetetlen paradigmákban különböző, de a kategorizálás, magyarázat és előrejelzés sikeressége által visszaigazolt mentális modulok alakulnak ki, melyek a környezet más-más ingereit minősítik relevánsnak, s ezért eltérő, de empirikusan egyaránt adekvát észleleti világokat konstruálnak. Ezért használja Kuhn a kognitív sikeresség plurális voltának demonstrálására az illúziók helyett az alaklélektan *többértelmű ábráit*. A mentális modul rendelkezik a fodori neurális modul előnyeivel: gyors, mivel szelektál, s csak a releváns ingerekkel foglalkozik. Kognitíve zárt, mivel az észlelést nem elméletek, hanem egy közösség gyakorlatán alapuló programozás befolyásolja, s így nem is változhat információk hatására, egyik pillanatról a másikra. Végül, a paradigmák veridikus (de nem uniform) módon reprezentálják a környezetet. Mivel tanult program, a neurális modultól eltérően, a paradigma *adaptív*, képes a változó környezethez (új kísérleti, mérési, megfigyelési adatok, új elmélet, fogalmi eszközök megjelenése) alkalmazkodni.

## II. SZEMANTIKA

Miután bizonyítottnak tekinti a tapasztalat objektivitását, Fodor elfogadhatónak látja a logikai pozitivisták kétnyelv-modelljét is. Moduláris elméletéből következik, hogy a megfigyelési terminusok jelentése közvetlenül, a tapasztalatok segítségével, a teoretikus terminusoké pedig a közös igazságfeltételekre visszavezetve definiálható, s így megoldhatatlan fordítási problémák sem támadhatnak. Fodor szemantikai kísérlete nem sikeres, de ismeretelméletileg reflektáltabb, mint a közös tapasztalat lehetőségét egyszerűen föltételező Kripke- és Putnam-féle merev deszignációs elmélet. Ez, ismeretelméletileg naiv módon,



abból indul ki, hogy a szavak referenciáját rögzítő keresztelési aktus elméletmentes igazságfeltételeket biztosító környezetben történik, ezért a rámutatás képes mindenki számára azonosítható entitást kijelölni. Ugyanúgy, mint a logikai pozitivistáknál és Fodornál, az elméleteket is mozgósító interpretálás a merev deszignáció relációjának megszilárdulása után történik, ezért, a föltevés szerint, a konceptuális és koncepcionális összetevő jól elkülönül.

Az azonos kultúrában élők köznapi tapasztalataiban fölbukkanó individuumokra alkalmazva, ez az elgondolás plauzibilis, de a tudományra aligha alkalmazható. Már az *individuumnevekkel* is komoly nehézségek vannak, hiszen a tudományban az egyedi entitásokat sem lehet elméletfüggetlenül azonosítani, lévén a legfontosabbak (a teoretikus entitások) megfigyelhetetlenek. Az igazán súlyos probléma azonban abból származik, hogy a tudomány nem egyedi entitásokról, hanem dolgok, események, folyamatok stb. *fajtaíró*ról szól. Putnam ebben sem lát külön problémát: bizonyos egyedeinek megmutatása és az „azonos természetű” predikátum segítségével kijelölhetőnek tekinti a fajtanevek extenzióját. A kuhni szemantika felől nézve, a definíció mindkét eleme elhibázott. A megfigyelhető individuális entitások osztenzív azonosítása – mint Wittgenstein és Quine megmutatta – függ a használt kategoriális kerettől, s ugyanez a helyzet az „azonos természetű” predikátum által föltételezett esszenciával is. Az aktuális paradigma jelöli ki a fajtadefinitív tulajdonságokat, s e készlet konszenzuálisan elfogadott és változó. Mivel az „azonos természetű” predikátumnak nincs *paradigmafüggetlen értelme*, a merev deszignáció elméletét akkor sem lehetne kiterjeszteni a fajtanevekre, ha az individuumnevek esetében problémátlan lenne.

Mint hogy az empirikus igazságfeltételek elméletfüggetlenségét nem ismeri el, Kuhn szemantikája relativizmust implikál. Kérdés azonban, hogy a paradigmák által kialakított kategorizáció teljes mértékben uralja-e az empiriát, az ingerek kizárólag az örökölt kategóriákban tudatosodhatnak-e. A nyelvről szóló fejezetben amellet is hozok föl érveket, hogy Kuhn jelentésemélete az igazságfeltétel-szemantika alkalmazhatatlansága ellenére sem billen át a Sapir-Whorf-féle nyelvi relativizmusba: ahogy a perzeptuális, úgy a nyelvi különbségek kialakulása is megelőzi az explicit elméleteket. Mire a nyelv a maga grammatikájával és lexikai kategóriáival színre lép, az alapvető rendezést, szelektálást már elvégezte egy mintakövetéssel kialakított, nem tudatos, „*nyelv előtti taxonómikus modul*”. Kuhn olyan intenzionális kategóriákról beszél, melyek nem öltenek testet szavakban, verbális általánosításokban, nem tudatosak, s már az állatok viselkedésében vagy az emberi észlelőrendszer működésében is kimutathatók. Eszerint a *konceptuális séma* nem azonos az elméletek által alkalmazott *nyelvi kerettel*, inkább a gyakorlatban kialakuló, „hogyan” jellegű diszpozíció.

Kuhn jelentésfelfogását ezért úgy értelmezem, mint a logikai pozitivisták empirista univerzalizmusa és a whorfi konceptualista relativizmus *közötti*, leginkább wittgensteiniánusnak tekinthető pozíciót: sem a nyers tapasztalat nem határozza meg a fogalmi struktúrát, sem a fogalmi struktúra nem determinálja a tapasztala-

tot. Mindkettő mögött a közösségnek a realitáshoz igazodó, de az által meg nem határozott, nyelvi és nem nyelvi *gyakorlata*, szokásokra, készségekre épülő mentális modulja áll, mely az ingerekből tudatos tapasztalatot, a tulajdonságokból fajtákat állít elő. A benyomások káoszának e diszpozíció általi perceptuális strukturálása s a fajták megnevezésére szolgáló lexikon kialakítása egyazon folyamat, melyben egyszerre formálódik az észleleti világ s annak nyelvi reprezentációja. Mint Kuhn mondja, a kettő *egyazon érme két oldala*.

A tapasztalatról és nyelvről szóló részek közös konklúziója, hogy az empirikus adatok objektivitása mellett felhozott érvek nem meggyőzőek, így az empirista észleléselemélet és szemantika fodori és putnami változata sem biztosíthatja a realizmust, racionalitást és közös nyelvet. A könyv második, „Társadalom” címet viselő fő része amellet érvel, hogy Kuhn ennek ellenére nem parttalan relativizmust implikál, irracionális vállalkozásként, hanem speciális normák szerint működő társadalmi intézményként írja le a tudományt.

### III. A TÁRSADALMI DIMENZIÓ

Első pillantásra nem világos, hogy az örökölt *ideál* helyett, a tudománytörténettel összhangban álló, társadalmi intézményként értelmezett *valóságos tudomány* középpontba állítása hogyan segít elkerülni a relativizmust. Ha a tudományfilozófia által végzett racionális rekonstrukció nem talál időtlen és univerzális ismeretelméleti és módszertani *normákat*, a különböző történeti és kulturális kontextusokkal kapcsolatba hozott tudomány történeti, szociológiai *leírása* még sokkal kevésbé lehet erre képes.

Ezen a ponton fontossá válnak a könyv első felének belátásai. Minthogy Kuhn nem számolta föl az empirikus-teoretikus dichotómiát, megmutatható, hogy a valóság általános empirikus nyomásának formájában jelentkező *minimálrealizmus* egész tudományfelfogásának alapvető eleme. A *valósághoz* nem közvetlenül, csak a benyomásokból a mindenkori fogalom- és műszerkészlet által létrehozott *episztemikus világ* formájában férünk hozzá. Mégis beszélhetünk realizmusról, mert a valóság metafizikai értelemben független a rá vonatkozó vélekedésektől, és a kapcsolat *kétirányú*: nemcsak a megismerő kategóriái formálják a beérkező ingereket, de ez utóbbiak is képesek kikényszeríteni a mentális és taxonómikus modulok átalakítását. E realizmus ismeretelméleti értelemben minimális: a visszajelzések korlátozzák a lehetséges perceptuális és lexikális világokat, de nem elég specifikáltak a valóság egyedül helyes megjelenítésének kiválasztásához. Azok az elméletek, melyek nem képesek konzisztens rendbe szervezni a relevánsnak minősülő empirikus adatokat, diszkreditálódnak, de arról nincs szó, hogy az egymást követő paradigmák egyre *igazabb képeket* adnának. Minthogy a valóság többféle (de nem akármilyen) tagolást-értelmezést enged meg, a minimálrealizmus a relativizmus korlátozásának *objektív föltétele*. Társadalmilag kon-

dicionált *szubjektív föltétele* pedig az, hogy Kuhn kantianizmusa darwiniánus, a paradigmákban megtestesülő kategória-rendszerek társadalmilag legitimált, tanult, így módosítható diszpozíciók. A minimálrealizmus és a tudomány társadalmi intézményként való működése együtt magyarázza a racionalitás megőrzését és a teljes relativizmus elkerülését. Lássuk, hogyan!

Könyvem fontos tézise, hogy Kuhn a magukat követőinek tekintő externalistákétól alapvetően különböző, *internalista tudásszociológiát* képviselt, melynek döntő felismerése, hogy a tudománynak két szintje van. *Transzparadigmatikus értékek és célok* vezérlik a paradigmákon átívelő működést, de ezeket az értékeket, a mindenkori ismeretszintnek megfelelően, az egyes történeti korszakokban eltérő példázatok, s a körülöttük kialakuló, *lokális-tentatív hagyományok* testesítik meg. E közös érték- és célrendszernek tudható be, hogy míg más intellektuális formákban a viták állandóan az alapvető vélekedésekre és módszerekre irányulnak, a (természet)tudományban a konszenzus rövid idő alatt mindig helyreáll. Az internalista tudásszociológiai olvasatra támaszkodva megmutatható, hogy a paradigmaváltás irracionalista interpretációja szövegszerűen és koncepcionálisan egyaránt tarthatatlan. A mindössze néhány alkalommal felbukkanó alakzatváltás metafora jelentőségének eltúlzása eredményezi azt a téves föltevést, hogy az anomáliákra a paradigmát fönntartó közösség *homogén* módon reagál, s a váltás egyetlen pillanat alatt lejátszódó *esemény*. Mivel Kuhn a paradigmát minta és hasonlóság által szervezett *tradícióként* írja le, a perceptuális és konceptuális mező *nem változhat hirtelen*, nem idézhetnek elő teoretikus belátások vagy információk az egész közösséget egyszerre érintő Gestalt-váltást.

A kutatást vezérlő értékek sokféleképp alkalmazhatók a konkrét helyzetekre, ezért az individuális különbségek döntő fontosságúak a tudomány dinamikája szempontjából. Ugyanakkor – a pszichologizáló félreértelmezéssel szemben – jelentőséget tulajdonítok azoknak a szöveghelyeknek, melyek a *lényegi feszültséget* nem a rendkívüli képességű individuális elmékből, hanem a tradicionalista és radikális újító tudósokat egyaránt tartalmazó *tudományos közösségből* származtatják. Az anomáliákra a közösség a kollektív normák lazításával, individuális alternatívák képzésével reagál. A *sikermaximalizáló és kockázatredukáló* kutatók különböző pillanatokban és eltérő radikalitású javaslatokkal lépnek színre, s ez a paradigma mutációinak kialakulásához, versengéshez, majd a tudomány általános értékeivel összhangban álló, pillanatnyilag elérhető leghatékonyabb megoldás kiválasztódásához és replikálódásához, azaz új paradigma létrejöttéhez vezet. A változás a paradigma biztosította kognitív eszközrendszer hatékonyságába vetett bizalom megrendülésének, helyreállításának vagy fokozatos lebomlásának, majd a kiválasztódott új minták köré szerveződő új kutatási tradíció felépülésének társadalmi *mikrofolyamataiból* áll össze. Irracionális ugrások, empirikus adatokkal összeegyeztethetetlen vagy logikai szabályokat sértő mozzanatok nem részei e folyamatnak, az eredmény pedig a réginél hatékonyabb kognitív eszközrendszer. A váltás az instrumentális hatékonyság (empirikus adatok ko-

herens rendszerbe szervezésének képessége, rejtvényazonosító és -fejtő kapacitás, magyarázó és előrejelző sikeresség) *összehasonlításán* alapul, bár empirikus konfirmációt vagy logikai következtetést alkalmazó *bizonyításról* nem, de indokokra hivatkozó *meggyőzésről* beszélhetünk, s ezért a váltás Kuhn szerint *racionális folyamat*.

A könyv záró része azt a kérdést vizsgálja, összekapcsolódnak-e a paradigma-váltások a tudomány hosszú távú fejlődésének iránnyal bíró *makrofolyamatává*. A standard interpretáció szerint nem, mert a periodikusan ismétlődő forradalmak szakadásokat eredményeznek a tudomány történeti szövedékében. Én úgy látom, Kuhn elképzelése ennek épp az ellenkezője: a tudomány története az egyre hatékonyabb kognitív eszközök kiválasztódásának hosszú távon kibontakozó, permanens forradalma. Ennek során, a kognitív eszközrendszer egyre komplexebb környezethez alkalmazkodik, azaz egyre nagyobb teljesítőképességűvé válik. A környezet komplexitásának növekedését új, korábban nem ismert jelenségek, elemek, folyamatok és hatások tudatosulása idézi elő, s a valóság e rejtett, egyre mélyebben fekvő tulajdonságai csak részletekbe menő, szisztematikus kutatásokkal, azaz az adaptációs folyamatot újra meg újra megszakító, *normál szakaszok közbeiktatásával* tárhatók föl. Noha a tények és megoldott problémák nem kumulálódnak, az ismétlődő normál szakaszok által föltárt, egyre komplexebb valósághoz való kognitív alkalmazkodás csak egyre artikuláltabb elméletekkel, kifinomultabb fogalmakkal, matematikai apparátussal, instrumentumokkal és diszciplináris tagoltsággal lehetséges. Ezért a valóságos tudomány fejlődése Kuhn szerint nem kontinuos és kumulatív, de nem is csapongó, hanem *egyirányú*.

Mindent egybevéve, jogosnak látszik, hogy Kuhn *fejlődésről*, mi több, *racionális* folyamatról beszél: az, hogy a tudományt társadalmi folyamatok tartják mozgásban, nem kizárja a racionalitását, hanem megmagyarázza. Az idealizált és valóságos tudomány közötti kuhni megkülönböztetés fő okaként a kognitív ágensről alkotott elképzelések különbségét jelöltem meg. Míg előbbi szerint a kognitív aktusok ágense a semleges és objektív empirikus adatokra és a tiszta logika abszolút szabályaira támaszkodó, racionális döntéseméleti alapon működő individuum, utóbbi szerint *a tudomány kognitív ágense* nem a társadalmi közegétől teljesen elkülönülő *individuum*, de nem is a kognitív folyamatokat monopolizáló *közösség*, hanem a kettő *intézményesen szabályozott együttműködése*. A speciációt és szelekciót eredményező egyirányú fejlődési folyamatot, a hagyományhű és hagyománytörő szakaszok ritmikus váltakozását a tudomány általános értékei vezérlik, s a szimbolikus tőke és kutatási lehetőségek elosztásának intézményes mechanizmusai által motivált individuumok tartják mozgásban. Ez a felismerés tette Kuhn számára szükségessé és lehetővé a „racionalitás”, „tapasztalat és elmélet”, „tudományos nyelv”, „elméletváltás” és „tudományfejlődés” fogalmainak, azaz a tudomány természetének a bevett nézetétől alapvetően eltérő értelmezését. A tudomány permanens értékei összemérhetetlen, de kognitív

hatékonyságukat tekintve összehasonlítható, tentatív hagyományokban konkretizálódnak, melyek a valóság egyre nagyobb részét értelmezhető világgá alakító elméleteket hoznak létre. Könyvem azt a nézetet képviseli, hogy a tradícióhű és tradíciótörő szakaszok váltakozását irányító permanens és univerzális értékek szerepének megvilágítása igazolja Kuhnnek azt a – könyvének címében is megjelenő – meggyőződését, hogy a paradigmaváltások mikro- és a tudomány hosszú távú fejlődésének makrofolyamata „tudományos” és „szerkezettel” bír, a tudomány egésze racionális vállalkozás.

## Laki János – A tudomány természete. Thomas Kuhn és a tudományfilozófia történeti fordulata

Kuhn újraértékelése az utóbbi évtizedben a magyar filozófiai élet középpontjába került. Mi sem bizonyítja ezt jobban, mint hogy az utóbbi tíz évben két habilitációs előadás is szólt Kuhnról: Miskolcon Forrai Gáboré, a Budapesti Műszaki és Gazdaságtudományi Egyetemen pedig Margitay Tihaméré; végül pedig itt van Laki János monográfiája, valamint a Fehér Márta és munkatársai által szerkesztett gyűjteményes kötet.

Kiindulásként azt szeretném hangsúlyozni, hogy mindez a magyar filozófiát érdekli és érinti a jelen pillanatban, ami sok szempontból – és nem érdektelen módon – önbezárkózásra utal. A konferenciákon alig vesznek részt nem filozófusok. Laki egyébként kiváló könyve, amely Kuhn saját idejét tekintve összhangban van a pszichológiai fejleményekkel (például Jerome Bruner és a perceptuális tanulás kutatás 1950-es évekbeli új szemléletének jelentőségével), a mai pszichológiára egyáltalán nem tér ki, nem követi, hogy mi történik azokkal a szakmai referenciákkal, amelyek Kuhn számára fontosak és megalapozóak voltak. Nem követi a hazai nem filozófiai kognitív vizsgáldást sem, amikor az Kuhnra reflektál.

Kuhnnek is több olvasata jelent meg, s e többféle olvasat számos szakmában jelen van. Laki kitűnő, egyszerre tiszta fogalmi elemzést adó és filológiaiag pontos könyve világosan érzékelteti, hogy az értelmezés attól függ, hogy melyik részre helyezük a hangsúlyt. A hangsúlyt pedig a tudósok mint hús-vér lények éppenséggel saját beágyazódásuknak, s a Kuhn által oly sokat hangsúlyozott paradigma-beállítódásuknak megfelelően hol oda, hol ide helyezik. Az egyik esetben a hangsúly a paradigmán és a normál tudományon van. Azon a tudományon, amelyik évszázadokon keresztül csak a részletek kidolgozásával foglalkozik, és ahol a mesterek, a tankönyvek és az intézmények által átörökített tudományművelés kánonjai uralkodnak. Ilyenkor Kuhn üzenete konzervatív. Azt üzeni, hogy el kell fogadnunk a rendet olyannak, amilyen, gondolkodásunkat pedig döntően befolyásolják ezek a rendet közvetítő episztémikus sémák, vagy kognitív mintázatok, alakok, Gestaltok. Ha viszont Kuhn könyvének másik oldalát, a címadó jelszó, a tudományos forradalom értelmezését vesszük, akkor Kuhn a radikális

változás és a szabadság szövetségese. Annak a felfogásnak a szövetségese, amely szerint értelmezéseink, – ahogyan Laki Kuhn és Wittgenstein terminológiáját visszaadja – *aspektusaink* gyökeresen meg tudnak változni.

Ha ebből a szempontból nézzük, akkor az 1960-as, majd az 1970-es évek számos szakmája számára Kuhn üzenete alapvetően forradalmi és felszabadító volt. Lelkesen olvasták azok az irányzatok, amelyek, miként arra Laki könyve nagyszerű szociológiai elemzésekben mutat rá, véletlen kontingenciák révén éppen gyökeres átalakulásban voltak. Ez az a kor, amelyben létrejön a molekuláris biológia, ami drámai változást eredményez a világszemléletünkben. Ez az a kor, amelynek vége felé kibontakozik a modern idegtudomány, ez az anatómia, az élettan, a szövettan, az evolúciós biológia, a kémia, a genetika stb. összekapcsolásával s az új számítástechnikai eszközök alkalmazásával az idegrendszer egészét vizsgáló rendkívül komplex diszciplína. A saját kutatási területemet tekintve ez az a kor, amelyben Chomsky munkássága révén egy forradalmian új szemlélet jelenik meg a nyelvről mint nyitott és ugyanakkor szabályok által irányított rendszerről, ahol a szabályok lényegéhez éppen a nyitottság fenntartása tartozik. Ennek egy átfogóbb üzenete az lesz, hogy a nyelv az ember különlegességének letéteményese. Ezzel párhuzamosan születik meg a kísérleti kognitív pszichológia is, amely a Kuhn által is oly sokat elemzett és Laki könyvében is oly sok formában visszatérő behaviorizmussal szemben felfedezi a belső embert. A belső ember úgy épül fel, hogy működésében majd központi szerepet játszanak olyan kérdések, amelyeket először valóban a Wittgenstein által sajátosan olvasott, Kuhn által értelmezett és Laki által elegánsan újra kiemelt korai alaklélektani kétértelműség-kutatás állít előtérbe. A kísérleti pszichológusok számára, ha az analógiát tekintjük a paradigmafogalom relativista és nem relativista értelmezései körüli vitákban, az alapkérdés mindig az lesz, hogy milyen mértékig, milyen sebességgel és milyen helyzetekben vagyunk meglévő rendszereink foglyai. Vajon hogyan tudunk túllépni a felülről lefelé történő információfeldolgozáson, vagyis azon, hogy sémáink határozzák meg, mit is észlelünk. Az a Piaget, akire – amint arra Laki is helyesen utal – Kuhn a gyermeki gondolkodás változásainak teoretikusaként hivatkozik, egyben egy másik kettősséget is megfogalmaz, amely másoknál, például Neissernél is a modern kognitív pszichológia alapproblémája lesz. Itt most szándékosan Piaget saját biologista kifejezéseit használom. Az emberi megismerés adaptív jellegű. Az adaptációnak azonban két pólusa van. Egyrészt alkalmazkodás a környezet feltételeihez (*akkomodáció*), ami megfelel a tudományos közösséghez való igazodásnak, a szociális meghatározottságnak a tudomány művelésében, amelyet Laki részletesen elemzett. Piaget számára a másik pólus azonban az *asszimiláció*: a beérkező adatok és beérkező hatások hozzáidomítása saját meglévő rendszerünkhöz, a felülről lefelé történő hatások. Ez az, amit Neisser majd a perceptuális ciklus problémájának nevez. Piaget számára mindez érinti a logika kiemelt státuszát is. Megismerőrendszerünkben a logika lesz az a környezetről való leválást lehetővé tevő rendszer, amely úgy biztosít

magasabb szintű adaptációt, hogy nem vagyunk állandóan becsapva saját akkomodációink és asszimilációink révén. A megismerés Piaget felfogása szerint állandó kompromisszum, fokozatos statisztikus közelítés. A közelítésben a logika azonban túllép a bizonyítás statisztikus módszerein. Ha Piaget-t kiegészítem Kuhnnal, azt kell mondanom, hogy az ismét csak Laki által részletesen elemzett és a percepció világával szembeállított, sajátos tudományos eljárásrendszer, a részekre bontás, az egyváltozós kísérletezés, a kísérletezés ellentetjeinek vizsgálata stb., egyáltalán a hipotézistesztelő modern tudományosság is azt eredményezi, hogy nem leszünk állandóan foglyai saját rendszerünk asszimilatív hatásainak, de attól még ezek léteznek.

Laki könyvének nagy érdeme, hogy Kuhnt nem az őt relativistaként értelmező hagyomány szerint állítja be. Ez utólag is igazolja azt, hogy miért voltunk sokan olyan lelkes Kuhn-olvasók. Mi az 1970-es évektől Kuhnról a szabadság szövetségeként s nem fogságunk megindokolásának és a misztikus maszlagfogalmak újrabevetésének szövetségeként tekintettük.

Laki munkájában Kuhn és a pszichológia illesztése szempontjából három sajátosan érdekes kritikus mozzanat van jelen. Az egyik a hosszan elemzett Kuhn–Fodor vita (85–108). Laki mint filosz kiválóan jár el, a pszichológus és kognitív kutató szempontjából azonban túl filoszra veszi a figurát. Itt nem ártott volna, ha olvas másoktól, például magyar irodalmat is. Történetileg is érdekes ugyanis, hogy Fodornál a kuhni felfogás kigúnyolása az illúziópéldákkal egy családott ember reakciója. Fodor azt hangsúlyozza, hogy érzeteink feldolgozása során kell, hogy előre rögzített rendszereink legyenek, Fodor részben ezt nevezi moduloknak. Olyan rendszerek ezek, amelyeket az elvárások stb. nem befolyásolnak. Ez éppenséggel kudarcainkban látszik jól – mondja Fodor. Például a Müller–Lyer-illúzió, akármennyit nézegetjük is, nem változik meg. Laki érve az, hogy Fodor itt sok szempontból csúsztat. Kuhn gondolatmenetében nem az illúziókról, hanem a kétértelmű ábrákról van szó, Wittgenstein kacsanyúl és Neckerkocka példái nyomán.

A Fodor-féle gondolatmenet igen részletes elemzése mellett Laki okos érveket hoz annak bemutatására, hogy Kuhn felfogása végső soron egy evolúciós naturalizmushoz vezet. Olyan nézethez, amely Kantot és Darwint kapcsolja össze (amint azt egyébként már a 19. században is sokan hitték, például Helmholtz is). A dolog logikai szerkezetét tekintve abban is igaza van, hogy Fodor moduljait evolúciósan rögzített rendszerekként mutatja be. Ebből a szempontból érdekes azonban követni azt – s itt nem ártott volna a szakmai továbbvitel –, ahogyan az utóbbi tíz évben, a valóságos evolúciós pszichológiai értelmezések megjelenése során Fodor mintegy megtagadja ezt az értelmezést. Pinkerrel, Cosmidessel, Toobyval és az evolúciós pszichológia más képviselőivel vitatkozva, visszahozza egy olyan rendszerbe, amely megretten attól, hogy a következtetési szabályok, vagy éppen a logika is csak egy esetleges újabb modul lennének. Fodor számára a radikális modularizmus megkérdőjelezhető, s nem elfogadható. Hasonlókép-



pen nem elfogadható – s ez már eredetileg is így volt, jóllehet csak később vet-ték észre – az evolúciós pszichológia adaptációs szempontjai révén. Tehát az a gondolatmenet, amely szerint – s itt mintegy vele szembe fordítva Fodor saját szavait használom – gondolkodásunk adaptív értelmezése azt jelenti, hogy jól já-runk, ha bizonyos reprezentációink, bizonyos gondolataink érvényesek a világra, azaz igazak. Az igazság maga – Fodor új olvasatában – nem kaphat haszonelvű értelmezést. Nem számon kérni akarom ezt Lakin, csak éreztetni, hogy mind-azok a problémák, amelyek Kuhn közel fél évszázaddal ezelőtti teljesítményé-nek relativista és nem relativista olvasataiban benne rejlenek, továbbgyűrűznek akkor, amikor a Laki könyvében is oly világosan elemzett episztémikus ismeret-elméleti individualizmus biológiai kantianus értelmezést kap.

Egy másik érdekes mozzanat a *kétértelműség*. Laki hosszan elemzi (101–105), hogy milyen nagy szerepet játszanak Kuhn érvelésének alakulásában a kétér-telműségi kutatások. Laki olvasatában Kuhn számára ez arra példa, hogy „míg az illúziók esetében van értelme egyetlen helyes, veridikus észlelésről beszélni, addig a többértelmű változatok esetében nincs” (101). Ebből kiindulva Kuhn eljut oda, hogy e látásmódok csak egy különleges, első személyű nézőpont szá-mára léteznek, bár ezt az első személyű nézőpontot Laki nem használja a dolog leírására. Azóta az igen kiterjedt csecsemőkutatásokból, például Csibra Gergely munkáiból, valamint majomkutatásokból tudjuk, hogy a kétértelműség látható következményekkel megjelenik beszámolni nem képes személyeknél is. Ez a kétértelműség valójában a Kuhn és Laki által sokat hivatkozott, a vizuális él-mény szerveződését meghatározó *figura-háttér* bontás elkerülhetetlen következmé-nye. Mindig valamit valami előtt látunk.

Ez egyben azt is mutatja, hogy milyen paradox helyzetben van a filozófus, amikor filológiai és fogalmi módszerekkel elemzi azokat a ténybeli tudásokat, amelyek morzsaként egy másik filozófusnál megjelennek. Amikor kacsanyula-akat vagy Necker-kockákat elemzünk, ahogyan azok Kuhn-nál és Wittgensteinnél megjelennek, állandóan felmerül a kérdés: na, és mit gondol ma erről a termé-szettudomány? Úgy, mint ahogyan akkor, az akkori természettudomány az ál-taluk hivatkozott Gestalt-pszichológia gondolta volna. Az aspektusok és az ész-lelés felülről lefelé hatásainak egész kérdésköre különleges kettősséget kapott. S itt a neurobiológiának számos önértelmezése is van, ha nem is a filozófusok általi, de filozófiai típusú önértelmezése. Ma az alaklélektanos korszaknál sokkal határozottabban különböztetjük meg a látás alacsonyabb és magasabb szintű fo-lyamatait. Számos agykérgi folyamatot is beleértve alacsonyabbakról beszélünk egészen addig, amíg az idegrendszer valamilyen értelemben felületeket számít ki a beérkező fénynyalákból. Ezekkel a felületekkel azonban két különböző dol-got kezd. Az egyik dolog valamelyest megfelel a relativista felfogásnak, a másik pedig az univerzalista és igazságelvű felfogásnak. Maga a látórendszer két útra oszlik, az egyiket, mivel a fali lebenyben folytatja legfontosabb komputációit,

dorzális útnak, a másikat viszont, mivel a halántéklebenyben végzi a legfontosabb komputációkat, ventrális útnak szoktuk hívni. E két utat „hol” és „mi” rendszernek, illetve „cselekvés-észlelés” rendszernek is nevezik. Lényeges eltérésük abban áll, hogy amíg a dorzális rendszer a cselekvéshez való hozzáférést képviseli, s énközpontú vonatkoztatási keretet ad, addig a ventrális rendszer tárgyközpontú leképezést használva a tárgyak mint tárgyak (mint *asztal*, mint *szék*) felismerését irányítja. Úgy tűnik, hogy a világhoz igazodó *tudni hogyan* jellegű tudások és a *tudni mit* jellegű tudások nem egyszerűen kétféle aspektus vagy kétféle attitűd kérdései, amelyeket szerencsésen váltogathatunk, hanem párhuzamos feldolgozások, melyek a főemlős és az ember látványkezelésének elengedhetetlen részei. Ez tulajdonképpen azt sugallja, hogy azon a Kuhn által elindított úton, ahol a tudományelmélethez egy természettudományos attitűdöt használó pszichológia adataiból szemezgetünk, tovább kell menni. Ma már nemcsak olyan adataink vannak, amelyek az új szemlélet szerint a felülről lefelé séma és fogalom-irányította asszimilációs feladatokat tükrözik, hanem olyanok is, amelyek valamelyest megfelelnek a hagyományos ismeretelméleti képelméletek veridikusészlelés-elvének.

Végül még egy pozitív mozzanattal zárnam az ismertetést. Amikor a paradigma és az észlelés sémafüggésének egymás mellé állítása jelenik meg Kuhn relativista olvasatában, valójában megfelelkezünk Kuhn figyelmes és gondos természettudósi olvasásáról. Laki nem követi el ezt a hibát. Miként a 104–105. oldalon részletesen elemzi, Kuhn tisztában van azzal, hogy az általa kiemelt relativista adatok nagyon rövid idők alatt érvényesülnek. A perceptuális tanulási hatások, a gyors kategóriába sorolás stb. néhány száz milliszekundumos, sőt, ma azt mondanánk, hogy néhány tucat milliszekundumos ablakban jelennek meg. Ezzel szemben a tudomány működése azért reflektív, mert van ideje szemlélődni. Ez az, amit Fodor majd úgy fogalmaz meg, hogy a Kuhn által is hivatkozott perceptuális új szemlélet kutatások nagy tévedése az volt, hogy az észlelés fogalmát szuboptimális körülmények között vizsgálta. Fodor számára tehát az észlelés igazi folyamatai az észlelés optimális körülményei között ismerhetők meg. A gyors felismerés időkorlátai és a reflexió kettőssége nem akármilyen problémát takar. Tudjuk ezt jól mai tudatelméletekből is. Daniel Dennett például, amikor a tudatosság narratív metateóriáját támasztja alá – bizonyára ő maga is tudja –, csal az időekkel, amennyiben a színes mozgáslátás jelenségét veszi kiinduló példának. Ennek az a lényege, hogy egy kritikus idővel egymás után vetített piros négyzet és zöld kör az elmozdulni látszó négyzet útján valahol félúton átvált pirosból zöldre. De a valóságos ingerhelyzetben két statikus kép volt. Ahhoz, hogy a pirosból zöldre átváltást megéljük, arra van szükség, hogy az emberi „elme” valahol megkonstruálja az a pályát, aminek a landoló pontja már nem piros, hanem zöld szín kell hogy legyen. Ez olyan, mintha egy történetet konstruálnánk, gondolja Dennett, és ugyanez történik akkor, amikor megkonstruáljuk magunkat,

a self, az én és a tudat mint a velünk történő események szerzőjének fogalmát és a belső színpadot is. Ha sztorik vannak, kell hogy legyen író is. Csakhogy jól tudjuk, hogy amíg a színes mozgáslátás lejátszódik, e folyamat egy néhány száz milliszekundumos ablakon belül történik. Dennett maga mondja, hogy itt a kritikus idő kb. 400-500 milliszekundum. Az a folyamat viszont, amelynek során az élettörténetünkben fokozatosan megkreáljuk magunkat mint szerető apát, mint odaadó tanárt, mint lelkiismeretes fiút stb., és kialakítjuk a személyiségünket, azt a kartéziánus színpadot, amin Dennett nem győz eleget gúnyolódni, évekig tart. Mindezzel csak azt akarom mondani, hogy Kuhn gondos olvasatának szükségessége is azt támasztja alá, hogy amikor a filozófusok – még a tudomány-filozófusok is – természettudományos eredményeket használnak fel, ezt óvatosan és kritikusan kell értékelnünk. Tudnunk kell, hogy ezek csak illusztrációk saját mondanivalójuk számára, és nem a ma érvényes tudományos faktualitások. Mutatják például azt, ahogy a tudomány sematizáltan, paradigmákon keresztül fejlődik.

## Kuhn recepciója, Laki rekonstrukciója

### I. MINIMÁLREALIZMUS, KORLÁTLAN RACIONALIZMUS

Laki János munkája sokat segít abban, hogy eloszlasson néhány általános tévképzetet Kuhnnal kapcsolatban. Komoly figyelmet szán annak, hogy a magát „javíthatatlan realistának” tekintő Kuhnt – szerintem nagyon inspiráló módon – elhelyezze a filozófiai realizmus térképén (113). Kulcsfontosságú, hogy Kuhn realizmusa *egyfajta* relativizmus, és Laki nemcsak a túl erős realista pozíciókat kritizálja, hanem felszólal a parttalan relativista álláspont ellen is (145). Szerintem helyesen a „minimálrealizmussal járó »korlátozott relativizmusnak«” tekinti Kuhn pozícióját (147). Ez ma is védhető álláspon: „A realitás és a tapasztalat között erősebb az összefüggés annál, hogy az előbbi pusztán lehetőségfeltétele lenne az utóbbinak, de annál gyöngébb, hogy e kapcsolat meghatározhatna egyetlen korrektnak tűnő leírást” (151).

Azonkívül, hogy nagyon érthető és világos Laki rekonstrukciója, külön érdekessége a szembenállás és a szembeállítás hangsúlyozása. Gyakran ír olyanokat, hogy általános az egyetértés a tudományfilozófusok között a tekintetben, hogy Kuhn irracionálisként írta le a tudományt (167). Egy alfejezet „A standard interpretáció szövegszerű megalapozatlansága” (169) mellett érvel. Laki hiszi, hogy kötetében „mód nyílik a Kuhn által leírt tudományos változás racionális értelmezésére. [...] Nem csak azt állítom, hogy ilyen értelmezés *lehetséges*, hanem azt is, hogy Kuhn *ténylegesen* inkább ezt a felfogást képviselte, s nem azt, amit a szokásos interpretáció tulajdonít neki” (178). Ezek az idézetek számomra szembenállást sugallnak egy hagyománnyal, és szembeállítják Laki olvasatát a „bevett nézettel”. Számomra ez kicsit meglepő, hiszen én egy olyan hagyományba szocializálódtam, ahol a kuhni meglátások triviálisak és senki sem lát bennük „árulást”, mint Weinberg (171). Mit is jelent ez pontosabban?

Laki a kritikusok és gyakran polemikus ellenfelek felől olvasva és velük szemben megvédvé Kuhnt bizonyos mértékig maga is polemikussá válik. Nem elsősorban azokra a hagyományokra épít, amelyek a kognitív munkamegosztás-

\* Munkámat a Bolyai posztdoktori ösztöndíj, az OTKA K 72598, a Magyar Állami Eötvös Ösztöndíj és a Collegium Budapest támogatta.

sal úgy próbálják megmenteni a tudomány racionalitását, hogy közben feladják a tudós racionalitásának auklérista eszméjét, és a modern – még csak formálódó, heurisztikákkal és nem dedukcióval dolgozó – kognitív tudomány modelljeit használják (lásd Kitcher, Longino vagy Solomon munkáit).<sup>1</sup> Ez a tradíció hasonló a Laki által körvonalazott Kuhnhoz – de Laki nem innen nézi Kuhnt, mint ahogy nem is abból az SSK / STS<sup>2</sup> hagyományból, amely valóban örököse Kuhn-nak, vagy akár Xiang Chen munkái felől (Anderson 2006). Holott „az azonos tanulási, kiképzési folyamaton átment emberek maguk válnak hasonlóvá egymáshoz” (149) – és igenis vannak olyan közösségek, amelyek nem Putnamon és Searle-ön, hanem Kuhnon, Flecken és Duhemen – vagy a tudásszociológia hagyományán – nőnek fel (Fehér–Békés 2005).

Ennek az érdekes nézőpontnak van egy apróbb, szerintem inkább negatív és egy fontosabb, határozottan pozitív folyománya. A negatív az, hogy Laki egy olyan hagyományból építkezik, ami Kuhnt általában elutasítja, trivialisálja és torzítja. Miközben itt megvívja harcát, és Kuhn kapcsán rámutat *ennek* a hagyománynak számos hibájára, valójában a hagyomány egyéb „dogmáit” elfogadja például Bloor és Feyerabend nézeteinek jellemzésekor. Hiszen – néhány szerencsétlen megjegyzésen kívül – David Bloor is az itt rekonstruált Kuhn-pozícióhoz nagyon közeli állásponton van, és álláspontja ugyancsak korlátozott relativizmussal írható le. Az utóbbi években többször is – görcsös? – kísérletet tett arra, hogy rámutasson: relativizmusa alapjában az abszolutizmus tagadása, és egyáltalán nem a parttalan relativizmus elfogadása. Bloor többször hangsúlyozta a természet szerepét a természet reprezentációjának kialakulásában és azt is kifejti, hogy miért *nem* kell ezzel egy szociológiai elemzés esetén foglalkozni (Bloor 1999; Kutrovátz–Zemplén 2008). Vagyis az a hagyomány, amelyet Laki Kuhn kapcsán korrigálni kíván, elfedi előle azt, hogy hasonló revízió járna Bloor-nak is. Sőt, az újabb értékelések alapján úgy tűnik, hogy akár a Feyerabend által erőltetett episztémikus pluralizmusnak is van helye a racionalitásfelfogások térképén (és nem mellette vagy a térkép másik oldalán, vö. Sunstein 2003). Ahogy Farrell monográfiájában megjegyzi: „bár úgy tűnik, Kuhn és Feyerabend számos, egymással diametrikus oppozícióban álló eszmét képvisel, valójában félúton találkoznak. Különösen ez a helyzet, amikor összehasonlítjuk nézeteiket a tudományos racionalitásról” (Farrell 2003. 163).

Laki nézőpont-választásának a határozottan pozitív következménye az, hogy egy sokkal argumentatívabb hagyomány eszközeivel mutatja be azt a Kuhnt, akit a tudomány-tanulmányok és a modern tudománytörténet-írás is hasonlóan

<sup>1</sup> Ezt az egész hagyományt Kuhn meglátása alapozza meg: „....Kuhn nemcsak elismerte létezését, de fontos funkciót is tulajdonított annak a ténynek, hogy *a tudományos közösség egyes tagjai nem azonos módon ítélik meg a helyzeteket*” (190).

<sup>2</sup> A tudományos tudás szociológiai elemzése (*Sociology of Scientific Knowledge*) és a tudomány és technika-tanulmányok (*Science and Technology Studies*) eltérő, de sok ponton kapcsolódó részterületei a tágran tudomány-tanulmányoknak nevezett hagyománynak.

lát, de már elszokott a filozófusok meggyőzésétől, legyintve és ironizálva azon, akik szerintük nem jól reagáltak az elmúlt fél évszázad legnagyobb szellemtörténeti kihívására. Ez a hagyomány elszokott a vitáktól más közösségek képviselőivel (és főleg attól, hogy az ő normáik szerint vitázzon), hiszen kialakította saját intézményi hátterét – nem a filozófiához kötődve, ahogyan még a 60-as, 70-es években tette, hanem elsősorban a történelem és szociológia tanszékek környékén. Ez a hagyomány kezdi problémának látni, hogy bár kutatók százai foglalkoznak a tudománnyal, alig kíváncsi a társadalom a véleményükre, amikor a tudománnyal kapcsolatos döntéseket készít elő. Ezt a normativitás iránti igényt azonban már nem a filozófiához kapcsolódva képzelik el (Collins–Evans 2002; Lynch 2006) – de hogy miként, arról e hagyományban nincs konszenzus. A szociologizáló tudománytörténet újra szeretne meghatározó lenni a tudománykép kialakításában (Goody et al. 2008), de csekélyek a lehetőségei, és nem bővelkedik eszközökben. Ebben a helyzetben válik különösen fontossá az a „hídverő” szerep, amit Laki könyve betölt – remélhetőleg segít abban, hogy újra egymásra találjanak a normativitás és a deskriptivitás kutatói, a filozófusok és történészek/szociológusok, egy olyan kultúrában, ahol a tudomány megértése és egyben megítélése a társadalomtudományok egyik legfontosabb feladata.

## II. PROBLEMATIKUS PONTOK

Azt hiszem, a fentiek alapján nem kétséges, hogy fontos, jól megírt kötetnek tartom Laki Kuhn-monográfiáját, és hogy legtöbb megállapításával mélységesen egyetértek. Éppen emiatt szeretnék néhány pontot kiemelni, ahol Kuhnmal kapcsolatos nézetem különbözik az övétől, illetve ahol Laki rekonstrukcióját problematikusnak tartom.

Az első problémám a „modul” fogalmával van. Azt még érteni vélem, amikor Kuhn arról beszél, hogy „van nyelv előtti taxonomikus modul, mellyel az állatok rendelkeznek” (130), hiszen ilyeneket a mai kognitív tudósok is mondanak. Azt azonban már nem értem, hogyan lehet igaz az, hogy „a lexikon (pontosabban, a tapasztalatok rendezését végző *mentális modul*, amelynek nem kell nyelvi-fogalmi formát öltenie) evolúciós termék, s mint ilyen, jelentős részben *valamennyi ember számára azonos*” (112), és ugyanakkor az is, hogy létezhetnek „Különböző mentális modulokat kialakító közösségek” (150). (Laki „Kuhn gyakorlással kialakított mentális modulja”-ként is hivatkozik a modulra, vö. 97.) Szerintem maga Kuhn sem konzisztens abban, hogy egy elkülönült modulról beszél, vagy valamilyen más, modulokat átívelő *top-down* folyamatról – ilyet a binokuláris rivalizáció kutatásakor ma már ismerünk (Logothetis 2004). Ennek tiszta kimondása, illetve a kuhni „modul”-konceptió kritikai vizsgálata szerencsésebb lenne a mostani tárgyalása során. (Olyan kérdések kritikai vizsgálatára gondolok, mint: A taxonomikus modul ugyanaz, mint a mentális? Hány modul is van?, Filoge-

netikailag mi humánspecifikus? Egy közösség kialakíthat modulokat? De akkor mit értünk „modul”-on? Stb.)

Hasonló tisztázásnak örültem volna Kuhn „anomália” fogalma kapcsán. Kuhn – és Laki is – rendre objektív tényként utal az anomáliára: „el kell ismerni, hogy *anomália* lépett fel” (17). Vagy: „Az anomália ellenben a paradigma adott területre érvényes tartalmi és/vagy instrumentális részének alapvető elhibázottságát, a fundamentális tudáselemek, metafizikai meggyőződések tarthatatlanságát, a használt fogalmak vagy módszerek alkalmatlanságát *jelzi*” (188). Az anomália ezekben az idézetekben *fellép és jelez*, de az, hogy valami az anomália státuszával rendelkezik, nyilvánvalóan társadalmi folyamatok eredménye – éppúgy, ahogy a természet reprezentációja is társadalmi folyamat eredménye. Ez azért fontos, mert a poszt-kuhniánus tudománytörténet ugyanúgy társadalmilag konstruálttá tette a kuhni tudományfejlődés elemeit, ahogy ez az elmélet társadalmilag konstruálttá tette a természet képét. A kuhni terminológiának ez a fajta „relativizálása” – ami nagyon jól látszik például Golinski összefoglaló tankönyvében (Golinski 1998) – nekem hiányzik Laki kötetéből, amelyben olyanokat olvasunk, hogy a „konszenzust aztán mindaddig fenntartják, míg egy adott paradigmában rejlő rejtvényfejtő potenciált ki nem aknázzák” (183–184). Kuhn még beszélhetett arról, hogy a tudósok *felismerik* (188), hogy egy paradigma nem működik, ma sokkal aktívabb szerepben látjuk a tudósokat, és az elemzések azt vizsgálják, hogy kik azok, akik *elhitetik* egy közösséggel, hogy egy paradigma már nem működik. Erre nagyon jó példa Kuhn kortársa, Chomsky (Gervain–Zemplén 2008), de akár Newton Descartes-ellenes kirohanásaira is utalhatunk, stb. Az anomália tényét valaki *el tudja fogadtatni* egy közösséggel – gyakran ugyanaz, akinek az arra adott megoldását is elfogadja egy közösség. Ez itt sem jelent parttalan relativizmust, hiszen itt is igaz az, hogy nem bárki bármit, de van aki valamit valamiként tud elfogadtatni. Emiatt ugyan jogosan állítja Laki, hogy a „megoldási kísérleteknek egyelőre »ellenálló rejtvény« és az »anomália« közötti határvonal nem definiálható episztemológiai fogalmakkal, s így nagyon bizonytalan” (189), eközben nem mondja ki, hogy szociológiai rekonstrukcióval egyértelművé tehető a határvonal és az átalakulás, s a határ korántsem bizonytalan.

Az utolsó problémám a kuhni evolúciófogalom rekonstrukciójára vonatkozik. Laki megjegyzi, hogy a „tudományfejlődés evolucionista felfogásának képviselőjeként Toulmint, esetleg Campbellt, de leginkább Poppert szokás megemlíteni, Kuhnt jóformán sosem” (193). Ehhez kapcsolódik a kötet egyik markáns tézise: „Úgy vélem, a nem racionális álláspont Kuhnnak tulajdonítása valójában az evolúció gondolatának mellőzéséből fakad” (194). Az állítás egyik fele mellett – vagyis, hogy Kuhn evolúciós gondolatát mellőzték – Laki nem érvel, és ez inkább igaz a Laki által felhasznált hagyományra, mint a tudománytörténet Kuhn-recepciójára. Itt nyilvánvalóan mennyiségi kérdésről van szó, így ettől tekintünk is el. A másik, sokkal fontosabb következmény az, hogy az evolúciónál értelmezett Kuhn „racionális”.

Az evolúciós racionalitás témaköréhez kapcsolódóan három kritikai megjegyzésem van: Laki szerintem pontatlanul rekonstruálja az evolúciós folyamatot, nem a késői Kuhn – részletesebben kifejtett – evolúciós nézeteit veszi alapul, és néhány ponton inkonzisztens. Kezdjük ezzel az utóbbival, ami igazából egy apróság. Egy lábjegyzetben, amikor a nagy tömegekben replikálódó, sikeres megoldásról beszél, megjegyzi, „Ez természetesen Dawkins mémelméletének megelőlegezése Kuhnnál” (199. 36. jegyzet). Mivel a mémek terjedésének sikeressége nem jelzője vagy mércéje racionalitásuknak, a lábjegyzet félrevezető – itt a sikerességnek szerintem két felfogása keveredik. Ha Kuhn a mémelméletet előlegezné meg, akkor evolúciós modelljét nem lehetne semmilyen értelmes racionalitásfelfogásként rekonstruálni. A mémelméletet éppen az olyan kulturálisan fennmaradó, de „irracionális” folyamatok magyarázatára lehet jól használni, mint a divatjelenségek, az öngyilkossági hullámok – akár Goethe Wertherének, akár Michael Jackson halálának hatására. Ez a szemlélet replikátor-központú, éppen azért, hogy a teleologikus kultúrafejlődési modelleket (amelyek például a kulturális minták fennmaradásához racionalitást kapcsolnak) cáfolja. A metafora racionalista értelmezése ezzel szemben valamilyen optimalizációs folyamatnak tekinti az evolúciót.

A szándék, hogy a sikerességnek valamilyen – optimalizációs folyamatként értelmezett – evolúciós magyarázatát adja, máshol is megjelenik. Laki kijelenti, hogy a „versengés eredménye a [...] leghatékonyabb [...] eszközök kiválasztódása. [...] A hatékonyság nem mechanikusan értendő, [...] ez inkább a használat során kialakuló *egyértétség arról*, hogyan érdemes a konkrét helyzetre alkalmazni a tudomány általános értékeit, [...] [ezek] ugyan nem hoznak egyértelmű eredményeket, de a versengés során mégis kiválasztódik a leghatékonyabb eszközrendszer” (195). Azonkívül, hogy igazán nem látom, milyen érvek alapján állítja Laki, hogy a „leghatékonyabb” eszközök választódnak ki (az optimalizációs felfogás általában világnézeti alapú, előfeltevésként és nem empiriára alapozva jelenik meg), szerintem ilyet Kuhn ennyire általánosan nem mond. A hatékonyság valóban nő, ahogy erről Kuhn a késői szövegeiben elég sokat ír a „speciáció” kapcsán. Laki ugyan ismer(tet)i ezeket a szövegeket, és nagyon szépen használja azok bizonyos részeit a *Tudományos forradalmak szerkezete* (a továbbiakban: TFSZ) problematikus pontjainak értelmezéséhez, azonban az itt először kifejtett speciációs, fajkeletkezési modell lokális hatékonyságnövelésről szól. Ezt ugyan lehet a tudomány fejlődéseként értelmezni, de Kuhn elég egyértelmű azon a téren, hogy a lokális speciáció globálisan a szakterületek elkülönülését hozza létre, és egyre inkább szétszabdálja a tudományt – még ha az új „fajok” látszólag interdiszciplinárisak is. E felfogás szerint az „evolúciós diagramokhoz hasonlóan a tudományok fejlődésének ábrája elágazó fához hasonlít. Egy »taxon« időbeli fejlődését a diakrón elemzés, a taxonok számát és elrendeződését *t* időben a szinkrón elemzés vizsgálja. A speciációra épülő modell így pl. párhuzamba állítja a diszciplinaképződést és a speciációt, valamint a vizsgálatuk



során felmerülő hasonló problémákat (pl. új taxon keletkezésének időpontja). Az új »faj« a közös lexikonnal rendelkező, egymással kommunikáló szaktudósok csoportja” (Zemplén 2001). Tömören: a lokális optimalizálási folyamat, amiről Kuhn ír, egyben a tudomány fragmentálódását is jelenti, míg ha jól értem, Laki globális optimalizációs folyamatnak tekinti a tudomány fejlődését.

A késői Kuhn többször is beszél arról, hogy új, nagy, evolúciós alapokra épülő köteten dolgozik (Kuhn 1990, 1993). A kötet kéziratának feldolgozása még várat magára, de a megjelent tanulmányok azt sejtetik, hogy a modell a tudomány radikális *horizontális* elkülönülését hangsúlyozza a TFSZ *vertikális* elkülönítésével szemben. Mint ilyen, szervesen kapcsolódik az újabb, a tudomány megosztottságát képviselő hagyományhoz (Dupré 1993; Galison–Stump 1996). És – folytatva az evolúciós párhuzamot – ahogyan a faj definíciójából sem hiányozhat, hogy annak egy példánya nem tud más fajok egyedeivel életképes utódokat létrehozni, úgy „a tudósközösségeknél is alapvető az egymás közötti kommunikáció képességén kívül a képtelenség más tudósközösségek *teljes* megértésére, így biztosítva az egyes közösségek izolációját és ehhez kapcsolódó fejlődését. Ez a taxonómiai vagy lexikonbeli különbség *szükséges* a fejlődéshez (a diverzitás nő, a hatékonyság nő), nem pedig *gátolja* azt...” (Zemplén 2001). Ez a felismerés – amely Kuhnt is elbizonytalanította – mintha teljesen kimaradna Laki rekonstrukciójából, aki a késői szövegeket is a korai Kuhn evolúciós felfogásának rekonstrukciójára használja.

Laki inkább valamilyen szupraindividuális szelekcióról beszél. Amikor felvázolja a kuhni evolúciós modell három szakaszát, a másodikról ezt írja: „A közösség versengési folyamatban, gyakorlati kipróbálás által kiválasztja az azonosított anomáliák kiküszöbölésére legalkalmasabb elméletet, fogalmi sémát vagy eljárást” (203). A közösség ilyen típusú racionalitása („a piac újraárazta a részvényt”) szerintem rekonstrukciós „műtermék”. Ez ma elfogadott (neo)darwini mechanizmusokkal nem modellezhető (Sterelny–Griffiths 1999), hiszen a fennmaradás lokális döntések alapján történik, még *mielőtt* meg lehetne találni a „legalkalmasabb” elméletet. A versengésben fennmaradó elmélet lesz a legsikeresebb (hiszen ezt támogatják leginkább) – de nem a végső sikeresség *miatt* választódik ki, hiszen „a tudósoknak semmilyen formában nem áll rendelkezésére a tapasztalataiban, mérési eredményeiben megjelenőn túli valóság” (103), így nem tudja az utólag legsikeresebb magyarázatot választani, csak a lokálisan leginkább annak tűnőt. Nem lehet teleologikus a szelekciós magyarázat, ha darwini akar lenni.

### III. ZÁRÓ MEGJEGYZÉSEK

Az elmúlt évtizedek legátfogóbb és egyben legvállalhatóbb hazai Kuhn-olvasata ugyanabban az évben jelent meg, amikor Cambridge Tudománytörténet és Tudományfilozófia Tanszéke konferenciát szervezett „Kuhn és a tudományos

tudás szociológiája” címmel. Ugyanezen az egyetemen – a tanszék összes oktatójának részvételével – kutatószemináriumot is tartottak 2006-ban a témában. Ezen az azóta elhunyt analitikus filozófus, Peter Lipton megjegyezte: A rivális elméletek nem hasonlíthatók össze „objektív” mérce alapján, ha másért nem, azért, mert van, amelyik újabb és már bizonyított, van, ami még csak ígéretes. A konstruktivista történetírás egyik legfontosabb alakja, Simon Schaffer és a komunitáriánus episztemológiát kutató Martin Kusch mosolyogva bólogatott. A szakma egyik vezető tanszékén a történészek, filozófusok és szociológusok nemcsak fontosnak tartották Kuhn és a részben belőle táplálkozó hagyomány vizsgálatát, hanem az alapkérdésekben is egyetértettek.

Egy nagyon hasonló Kuhn képe bontakozik ki Laki kötetéből is – a különbségek csak árnyalatnyiak. Összefoglalóan: Laki a „korai” és „középső” Kuhnból ad nagyon erős olvasatot, a „késői” Kuhnt is leginkább ehhez használja. Így kötete számomra leginkább a TFSZ Kuhnjának nagyszerű rekonstrukciója.

Laki Kuhn kortárs kritikusa is több figyelmet fordít, mint kortárs támogatóira. Így, bár elmarasztalja a kritikusokat, csak korlátozottan épít az analitikus filozófia történetét radikálisan átíró közösség<sup>3</sup> eredményeire, amely az utóbbi évtizedekben elkezdte „historizálni” a tudományfilozófiát. Nem a konstruktivista szerzőkhöz fordul, hanem az analitikus hagyomány ellenvetéseit tárgyalja, és számos tévképzetet oszlat el, amely mindmáig jellemzi Kuhn magyarországi recepcióját.

Számomra mint tudománytörténész számára a könyv éppen e jellemzői miatt egy rég hiányzó hidat ver a mai magyar filozófia felől a modern szemléletű tudománytörténet-írás irányába. Erre égető szüksége van legalább az egyik szakmának. Remélem, hogy kötete katalizálja azt a szociológiai folyamatot, amely elfogadtatja Kuhnról, hogy „a tudomány változását nem artikulálatlan *pszichológiai eseményként*, hanem tagolt *szociológiai folyamatként* írja le” (179), és segíti a tudomány fejlődésének minél mélyebb megértését, hozzájárulva a szakterületek ismételt közeledéséhez.

## IRODALOM

- Andersen, Hanne – Peter Barker – Xiang Chen 2006. *The cognitive structure of scientific revolutions*. Cambridge, Cambridge University Press.
- Bloor, David 1999. Anti-Latour. *Studies in History and Philosophy of Science*. 30/1. 81–112.
- Collins, H. M. – Robert Evans 2002. The Third Wave of Science Studies: Studies of Expertise and Experience. *Social Studies of Science*. 32/2. 235–296.
- Dupré, John 1993. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA, Harvard University Press.

<sup>3</sup> A HOPOS-hoz (The International Society for the History of Philosophy of Science) kötődő kutatók köre.

- Farrell, Robert P. 2003. *Feyerabend and scientific values: tightrope-walking rationality, Boston studies in the philosophy of science*. Dordrecht, Kluwer.
- Fehér Márta – Békés Vera (szerk.) 2005. *Tudásszociológia szöveggyűjtemény*. Budapest, Typotex.
- Galison, Peter – David J. Stump 1996. *The Disunity of Science: Boundaries, Contexts, and Power, Writing Science*. Stanford, Calif., Stanford University Press.
- Gervain Judit – Zemplén Á. Gábor 2008. Miért (nem) empirikus tudomány a generatív nyelvészet? In Gervain Judit – Pléh Csaba (szerk.) *Láthatatlan nyelv (tanulmánykötet a Láthatatlan Kollégium diákjainak és tanárainak munkáiból)*. Budapest, Gondolat. 134–169.
- Golinski, Jan 1998. *Making Natural Knowledge: Constructivism and the History of Science*. Cambridge History of Science. Cambridge, Cambridge University Press.
- Goody, Graeme – John M. Lynch – Kenneth G. Wilson – Constance K. Barsky 2008. Does Science Education Need the History of Science? *Isis*. 99. 322–330.
- Kuhn, Thomas S. 1990. The Road Since Structure. *PSA*. 2. 3–13.
- Kuhn, Thomas S. 1993. Afterwords. In Paul Horwich (szerk.) *World Changes: Thomas Kuhn and the Nature of Science*. Cambridge, MIT Press. 311–341.
- Kutrovácz Gábor – Zemplén Á. Gábor 2008. A Bloor-Latour vita – Egy tudományos vita érveléstudományi vizsgálata. In Gervain Judit – Pléh Csaba (szerk.) *Láthatatlan nyelv (tanulmánykötet a Láthatatlan Kollégium diákjainak és tanárainak munkáiból)*. Budapest, Gondolat. 231–259.
- Logothetis, N. K. 2004. A látás: ablak a tudatra. In Pléh Csaba – Boross Ottilia (szerk.) *Bevezetés a pszichológiába*. Budapest, Osiris. 266–275.
- Lynch, Michael 2006. Expert Metascientists. *Social Studies of Science*. 36/6. 867–868.
- Sterelny, Kim – Paul E. Griffiths 1999. *Sex and Death: An Introduction to Philosophy of Biology*. Chicago, University of Chicago Press.
- Sunstein, Cass R. 2003. *Why societies need dissent*. Cambridge, Mass., Harvard University Press.
- Zemplén Á. Gábor 2001. A naturalizálás diszkrét bája. In Kampis György – Ropolyi László (szerk.) *Evolúció és megismerés*. Budapest, Typotex. 285–297.

## Laki, Kuhn és a realizmus

Laki János kiváló monográfiát írt Kuhn-ról. Kuhn igazából a konkrét történeti példák elemzésekor van elemében, de a példákból levont általános megállapításai gyakran homályosak és többféleképpen is értelmezhetők. Laki tökéletesen megbízható interpretátor. Belát a bizonytalan megfogalmazások mögé, s nagyon jól érti, hogy mi is az, amit Kuhn voltaképpen mondani akart. A könyvben nem találtam olyan pontot, ahol ne volna meggyőző, viszont számos olyan pontot találtam, ahol új és izgalmas dolgokat mond Kuhn-ról. Úgy hiszem, hogy a Kuhn-interpretáció műfajában nem is érdemes még egy monográfiát írni: ami érdekes, védhető, és Laki nem írta meg, abból nem futja egy másik monográfiára.

De a könyv nem csupán Kuhn-interpretáció, hanem filozófiai teljesítmény is. Lakinak van történeti rálátása Kuhn munkásságára, és persze hivatásos filozófus, így olyan módon tudja kifejtetni a kuhni tudományfilozófiát, ahogy arra maga Kuhn nem volt képes. Így a könyvből élénk táruló tudományképek részben Laki a szerzője. Ha az interpretátor maga is szerzővé lép elő, az azzal a veszéllyel jár, hogy interpretátorként megbízhatatlanná válik, mivel annak érdekében, hogy saját nézeteit és a szerző nézeteit integrálni tudja, meg kell hamisítania a szerzőt. Laki elkerüli ezt a veszélyt. Szerintem az teszi erre képtessé, hogy maga is kuhniánus. Erről számos dolog árulkodik. Egyrészt, legfeljebb gondatlan megfogalmazásaiért rója meg olykor-olykor Kuhnt. Másrészt, a Kuhn és bírálói közötti csatákban rendre Kuhn nyer. Szó sincs arról, hogy Laki elfogultan, kellő elemzés nélkül utasítaná vissza a bírálatokat. De tény, ami tény: mindig Kuhn nyer. Harmadrészt, a Kuhnt érő bírálatokra válaszul Laki sokszor olyan eredeti érveket fogalmaz meg, amelyeket aligha lett volna türelme alaposan kimunkálni, ha nem gondolta volna úgy, hogy Kuhn álláspontja helyes, és érdemes megvédeni.

Ebben az írásban Lakinak a realizmussal kapcsolatos felfogását szeretném bírálni. A kérdéskör jóval nagyobb annál, hogy körbejárhassam, ezért mindössze három kritikai megjegyzést fogok tenni. Egy: Laki félreérti azt a realista álláspontot, amely a kuhni minimálrealizmus alternatívája. Kettő: Laki érvei hatástalanok ez ellen az álláspont ellen. Három: Laki érvei az igazság korrespondenciaelmélete ellen szintén nem meggyőzők.

Az első problémához messziről kell nekifutnom. Hogyan referálnak terminusaink a valóságra, és hogyan tesznek szert kijelentéseink igazságfeltételekre? E kérdések megválaszolásakor a logikai pozitivisták abból indultak ki, hogy velünk született ideák és intellektuális intuíció híján a valósághoz nem férhetünk hozzá másként, mint érzéki tapasztalataink révén. A tapasztalat, a logikai pozitivisták terminológiájában a „megfigyelés”, jelenti azt a pontot, ahol nyelv és valóság összeér. A nyelvnek van egy olyan rétege, a megfigyelési nyelv, melynek terminusai közvetlenül a tapasztalatokra vonatkoznak. A többi terminus, az „elméleti terminusok”, pedig azáltal tesznek szert referenciára, hogy logikai kapcsolatban vannak a megfigyelési terminusokkal. Kiderült azonban, hogy e logikai kapcsolatok nem elégségesek a referencia rögzítésére, valami másra is szükség van. Ezt a más-t a logikai pozitivisták abban vélték megtalálni, hogy az elmélet axiómái részben meghatározzák a bennük szereplő terminusok jelentését, s így az elméleti terminusok referenciáját az axiómák és a megfigyelési terminusokkal való logikai kapcsolatok együttesen rögzítik.

Ekkor lép a színre Kuhn, aki észreveszi, hogy a „valami más is kell” problémája nem kizárólag az elméleti terminusok kapcsán merül fel. Nincs tiszta megfigyelési nyelv, mivel nincs tiszta tapasztalat. Az érzékszerveinket érő ingerekhez önmagukban nem férünk hozzá, ezek a tudat szintje alatt vannak, s ezért önmagukban nem is használhatók fel arra, hogy terminusainknak jelentést adjunk. Valami másra is szükség van: arra, hogy az észlelés elvégezze a maga szelektáló, kategorizáló, strukturáló munkáját, hogy megnevezhető entitásokká gyúrja át a pusztá ingereket. Valamilyen „idegi programozás” (Kuhn) vagy egy „mentális modul kialakítása” (Laki) szükséges ahhoz, hogy olyan entításokat kaphassunk, melyek szavaink referenciájául szolgálhatnak.

Kuhn következő állítása az, hogy ugyanilyen módon adhatunk számot az elméleti terminusok referenciájáról. Ahogy a kisgyermek megtanul labdát látni, s megtanulja ezt összekapcsolni a „labda” szóval, a tudóspalánta is megtanul ingát látni, tömeget látni, erőt látni, gyorsulást látni. Vagyis az elméleti terminusok referenciája ugyanúgy az észlelés révén adott, mint a megfigyelési terminusoké. Ebből egyrészt az következik, hogy szükségtelen azzal bajlódni, hogy az „inga”, „tömeg”, „erő”, „gyorsulás” szavakat megfigyelési terminusok révén határozzuk meg, másrészt pedig az, hogy valójában nincs értelme a megfigyelési és elméleti terminusok megkülönböztetésének, hiszen referenciájukat ugyanúgy rögzítjük („elméletterheltség”).

A referencia tehát előfeltételez egy észlelési képességet. Az észlelési képesség kialakulásának megvannak a maga feltételei. E feltételek sokfélék: biológiaiak, ökológiaiak, szociálisak, kulturálisak. A tudományos kifejezések esetében elengedhetetlenek a kulturális feltételek. Mindenki körül vannak erő hatására gyorsuló testek, de erőt és tömeget csak az tanul meg észlelni, akit fizikára tanítanak. Azt gondolhatnánk, a tudomány fejlődése abban áll, hogy a régi entításokkal kapcsolatban új belátásokra teszünk szert, illetve megtanulunk a

valóságban új entitásokat észrevenni, s ezekről új dolgokat mondani. Csakhogy, mondja Kuhn, vannak olyan tudományos változások („forradalmak”), melyekben nem ez, vagyis nem egyszerű hozzáadás történik, hanem a radikálisan átalakul az a mentális modul, amely a nyers ingereket megnevezhető entitásokká formálja. Ha kialakítjuk azt a mentális modult, melynek révén newtoni erőt észlelünk, akkor ezzel immár nem észlelhetjük az arisztotelészi erőt. Nem többet vagy jobban látunk, hanem mást látunk. S mivel mást látunk, másra is referálunk („referenciaváltozás”). Azokat az entitásokat, melyek korábban szavaink referenciájával szolgáltak, új entitásokra cseréljük le (ez az „összemérhetetlenség” egyik aspektusa).

Ezen a ponton pedig élesen felmerül a realizmus kérdése. Hol vannak az objektív, tudatunktól független entitások, amelyekre referálunk? Hol vannak az objektív, tudatunktól független tények, amelyek egyeznek vagy nem egyeznek kijelentéseink igazságfeltételeivel, s így kijelentéseinket igazzá vagy hamissá teszik? Eléri-e a nyelv magát valóságot? Sokak szerint Kuhn, bár ezt vonakodik elismerni, voltaképpen markáns antirealista álláspontot képvisel: tagadja, hogy objektív, tudatunktól független entitásokra referálnánk, hogy kijelentéseink igazsága objektív, tudatunktól független tényeken múljon. Laki szerint helyesebb azt mondanunk, hogy Kuhn minimálistalista. Az objektivitást nem feladja, hanem újféléképpen értelmezi.

„Kuhn redukálja az ontológiai függetlenséget: van független, metafizikai értelemben adott *realitás*, de ez számos (vagy számtalan) nem független, konstituált *világ* létrehozását teszi lehetővé” (114). Azok az entitások, amelyekre referálunk, és azok a tények, amelyek kijelentéseinket igazzá teszik, a mentális moduljaink által preformált világhoz tartoznak, nem pedig a mentális moduljainktól független realitáshoz. Ez azonban nem jelent antirealizmust, mert az általunk konstituált világok mégiscsak a tudatunktól független valóság ellenőrzése alatt állnak: „a valóságnak van egy bizonyos keménysége, mely nem tesz lehetővé bármilyen kategoriális rendezést” (uo). Bizonyos világokat a valóság elutasít, s ez abban nyilvánul meg, hogy a világ leírhatatlannak bizonyul, azaz olyan empirikus problémáink támadnak, melyek makacsul ellenállnak megoldási kísérleteinknek. E problémákban maga a realitás jelentkezik.

[A] természet közreműködik ugyan az elméletek kialakításában, az elvetésükről és átalakításukról szóló döntésben, de csak a gyakorlati alkalmazhatóságban vagy alkalmazhatatlanságban megmutatkozó általános empirikus nyomásként jelentkezik: számos taxonómiát megenged, de bizonyos kategorizálási kísérleteknek ellenáll. A realitás meghatározó ereje nem akkora, hogy egyértelműen dönthetne szavaink referenciájáról. Megakadályozza, hogy rá vonatkozó elméleteink bármilyenek legyenek (147).

A realitás tehát igenis szerepet játszik. De nem játszik más szerepet, mint hogy általános empirikus nyomást gyakorol. Ezért csak *minimálrealizmus* ez.

S most lássunk egy másfajta, nem minimális realistát. Ez a realista elfogadja, hogy a valósághoz nem férhetünk hozzá máshogy, mint érzéki tapasztalataink révén. Azt is elfogadja, hogy a tapasztalat nem elégséges a szavak referenciájának rögzítéséhez. De azt a választ, melyet a logikai pozitivisták vagy Kuhn és Laki ad a „valami más is kell” problémájára, nem tartja kielégítőnek. Nem mintha tagadná, hogy valami olyasmi, mint a logikai pozitivisták logikai kapcsolatai, elméleti axiómái vagy Kuhn és Laki mentális moduljai szerepet játszanának a referencia rögzítésében. Valami ilyesmire mindenképpen szükség van, hiszen egy pusztá hangsorozat önmagában nem bír referenciával: csak akkor bír referenciával, ha többé-kevésbé szabályos módon használjuk – a logikai pozitivisták, Kuhn és Laki pedig a nyelvhasználat azon mozzanatait igyekeznek feltérképezni, melyeknek szerepük van a referencia rögzítésében. A realista úgy gondolja azonban, hogy a referencia nem *csak* attól függ, hogyan használjuk az illető kifejezést. Attól is függ, hogy mi van, nem a mentális modulok által preformált világban, hanem magában a realitásban. Hadd tegyem ezt szemléletesebbé egy metaforával. A valóság és köztünk ott van az észlelés fátyla, ami azon túl van, ahhoz nincs közvetlen hozzáférésünk. A logikai pozitivisták és Kuhn kissé másként gondolkodnak e fátyolról: a logikai pozitivisták szerint adott, Kuhn szerint részben mi szőjük. Egyben azonban egyetértenek: a szavak referenciájának meghatározásában nem játszhat közre más, csak a fátyol, s ami a fátyol innenső oldalán van: pszichológiai rutinjaink, definícióink, elméleteink, nyelvhasználati szabályaink stb. A realista szerint az is számít, ami a fátylon túl van.

Miért gondolja ezt a realista? Abból indultunk ki, hogy a valósághoz csak érzéki tapasztalatainkon keresztül férhetünk hozzá. Ezt a realista úgy érti, hogy a valósághoz igenis hozzáférünk. Mivel hozzáférésünk közvetett, természetesen fennáll a veszély, hogy tévedünk vagy, rosszabb esetben, teljesen értelmetlenül beszélünk. Kockázat nélkül nincs igazság és nincs referencia. Ha viszont a referencia meghatározásában nincs szerepe maguknak a dolgoknak, hanem csak tapasztalatainknak, nyelvi szokásainknak és kognitív műveleteinknek, akkor a referencia nem szavaink és a dolgok között áll fenn, hanem szavaink és a fátylon lévő mintázatok – a „világ” dolgai, ahogy Laki mondaná – között. A realista szerint ténylegesen nem így gondolkodunk a referenciáról. Ennek megmutatására Kripke és Putnam érveit használja. Ezek az érvek úgy festenek, hogy veszünk egy episztemikusan hátrányos helyzetű közösséget (Putnamnál: akik nem tudnának különbséget tenni  $H_2O$  és XYZ között), amely egy bizonyos fátylon keresztül szemléli a valóságot, mely fátyol a mi látásunkat szerencsésen nem korlátozza. Az ilyen esetekben, mondja Kripke és Putnam, az episztemikusan hátrányos helyzetű közösség szavait olyan módon értelmezzük, hogy magukra a fátylon túli dolgokra referálnak (a  $H_2O$ -ra), nem pedig a fátylon kirajzolódó mintázatokra (színtelen, szagtalan, íztelen folyadék). Ez az interpretáció szoká-

sos módja. Akkor saját szavaink esetében miért tennénk kivételt, saját szavaink miért csupán a fátyol mintázataira referálnának?

Hadd fogalmazzam meg, immár metafora nélkül, hogy miben áll a különbség egyfelől a realizmus, másfelől a logikai pozitivizmus, Kuhn és Laki között. Az utóbbiak szerint a valóságnak csupán olyan vonásai képesek befolyásolni szavaink referenciáját és mondataink igazságfeltételeit, amelyeket valamilyen értelemben ismerünk, amelyeknek felismerésére észlelési rutinjaink vagy fogalmi kritériumaink vannak. A realizmus szerint a referenciát a valóság olyan vonásai is befolyásolják, amelyek teljesen ismeretlenek számunkra. Ez a fajta realizmus tehát abból a doktrínából táplálkozik, melyet más összefüggésben szemantikai externalizmusnak szokás nevezni.

Amikor azonban Laki a realizmussal vitatkozik, az imént vázolt képet kiegészíti két doktrínával:

1. A valóságnak van egy adott, igazi metafizikai szerkezete.
2. A szavak csak e szerkezet elemeire referálhatnak, akárhogy is használják a beszélők a szavakat (azaz: akármit is gondolnak arról, amire szavaik referálnak).

A realista álláspontot a következőképpen jellemzi:

Függetlenül attól, hogy mit tudunk, miben van igazunk vagy miben tévedünk, a világ valamilyenként meghatározott, s ez a tőlünk független külső tény externálisan meghatározza a szavak referenciáját: vonatkozó definícióink nem determinálják a világot, a természeti fajták a rájuk vonatkozó tudásunktól függetlenül, metafizikailag kötöttek azok, amik. S megfordítva: az, hogy a fajták metafizikailag adva vannak a világban, nem szükségképpen tükröződik szavaink szándékolt jelentésében. A természeti fajtákat megnevező szavak esetében mintegy elválnak egymástól a nyelv és a gondolkodás: a mintákkal való pragmatikus kapcsolat eredményeképpen e nevek magával a fajtával (minden dologgal, amely „ugyanolyan természetű, mint ez”) kerülnek kapcsolatba (136).

A feltevés szerint, aki például „denevért” mond, az állatvilág egy természeti fajtájáról beszél, melynek minden egyedét jellemzi bizonyos lényegi tulajdonság. Bármit gondoljon vagy tudjon, illetve ne gondoljon vagy ne tudjon is az illető e közös tulajdonságról, a denevérek csoportja a világban ettől függetlenül metafizikailag van meghatározva, s aki kimondja ezt a szót, erre az adott fajtára referál, akkor is, ha azt gondolja, hogy ugyanolyan állatok tartoznak ide, mint amilyenek a gólyák (134).



Vagyis: ha egyszer a denevér természeti fajta, a denevér-vagy-gólya pedig nem az, akkor a „denevér” szóval csak a denevérekre lehet referálni, a denevér-vagy-gólyákra nem, akármit is mond az illető a „denevér” szó felhasználásával. Ugyan miért fogadná el ezt a realista?

Laki benyomásom szerint olyan módon értelmezi a realistát, mint akinek célja annak a megakadályozása, hogy a tudomány forradalmi változásai során megváltozhasson a terminusok referenciája. Egyértelmű, hogy (1) és (2) elfogadásával ennek gátját lehetne vetni. Ha ezeket elfogadjuk, terminusainkkal csak bizonyos dolgokra referálhatunk, s bármit is mondunk e terminusokkal, e dolgokra referálunk. Hiába mond Einstein a tömeggel kapcsolatban olyasmit, ami a newtoniánus számára még csak nem is komolyan vehető elméleti lehetőség, ugyanarra referálnak. A valóság, a maga szerkezetével, tökéletes kontrollal bír a referencia felett, és megakadályozza annak megváltozását.

Úgy gondolom, a szemantikai externalizmusból fakadó realizmus ilyen értelmezése történetileg és doktrinálisan is hibás. Történetileg azért, mert Kripke és Putnam nem Kuhnra reagálnak. Kripke tudomást sem vesz Kuhnról, Putnam pedig elsősorban a logikai pozitivista örökség felszámolásán munkálkodik – bár kétséggkívül látja, hogy az ő nézőpontjából Kuhn és a logikai pozitivisták közös nevezőn vannak, amennyiben a referencia rögzítését kizárólag nyelvhasználati tényekre bízják. Doktrinálisan azért, mert a szemantikai externalizmusból nem következik sem (1), sem (2). Erre éppen Putnam a példa, aki belső realista korszakában változatlanul fenntartja a szemantikai externalizmust, viszont egyértelműen tagadja (1)-t és (2)-t. (1)-et „metafizikai realizmusnak” nevezi – Laki éppen tőle idézi a metafizikai realizmus jellemzését (113–114) –, (2)-t pedig „mágikus referenciaelméletnek”.<sup>1</sup>

De vegyünk inkább egy Kripkéhez hasonló metafizikai realistát, aki szerint az individuumok azonossága individuális lényegükön, a természeti fajták azonossága pedig esszenciális tulajdonságaikon múlik, mely individuális lényegeket és esszenciális tulajdonságokat metafizikailag adottak a dolgok rendjében, és teljesen függetlenek attól, hogy mit tartunk felőlük. (2)-t azonban egy ilyen metafizikai realista sem fogadná el.<sup>2</sup> Tekintsünk valakit, aki „denevérnek” nevezi a gólyákat is, szisztematikusan, és tökéletes látási viszonyok között is, ráadásul a „denevér” szóval nevezett fajt természeti fajtának tartja, mégpedig a kripkei értelemben. Tegyük fel továbbá, hogy ha felhívjuk a figyelmét a gólyák és a denevérek közti szembetűnő különbségekre, akkor készséggel elismeri ezek létezését, de részletesen elmagyarázza, hogy a biológiai taxonómia szempontjából

<sup>1</sup> Hilary Putnam, *Reason, Truth and History*. Cambridge, Cambridge University Press. 1981. 1–3. fejezet.

<sup>2</sup> Ahhoz, hogy egy szó referenciáját az individuális lényeg vagy az esszenciális tulajdonságok rögzítsék, minimálisan arra van szükség, hogy az illető szót tulajdonnévként, illetve természeti fajta névként használjuk. (2)-t úgy kell értenünk, hogy nem tagadja ezt a nyilvánvaló tényét.

ezek a különbségek jelentéktelenek, és figyelmen kívül hagyandók. A metafizikai realista ekkor is azt mondaná, hogy a „denevér” szó a denevérekre (vagy a gólyákra?) referál? Nem inkább azt mondaná, hogy a denevér-vagy-gólyákra referál, csak éppen az illető tévesen véli azt biológiai fajnak? Márpedig, ha elfogadná (2)-t, az előbbit kellene mondania. Röviden: (2) egyszerűen implauzibilis, s nem világos, hogy miért fogadná el bárki is.

S ezzel rá is térek a második kritikai megjegyzésre. Laki érvei a realizmus ellen előfeltételezik, hogy (2) része a realista álláspontnak. Ha jól értem, két ilyen érve van. Az első egy Kuhnból merített történeti érv (137–139). Eszerint a tudományos klasszifikáció során felhasznált kritériumok történetileg változó, különböző tudósközösség más-más hasonlóságokat ítélnék az osztályozás szempontjából relevánsnak. Hajdanában a kémiai klasszifikációknál alapvetőnek tekintették a halmazállapotot, míg ma a molekuláris szerkezetet tekintjük alapvetőnek. A második érv ennek az elméleti változata (140–145): a hasonlóság logikája kizárja, hogy az a követelmény, miszerint az azonos fajtakba tartozó példányoknak hasonlóknak kell lenniük, önmagában meghatározhassa a fajtak határait. Számptalan hasonlóság van ugyanis, továbbá a hasonlóság fokozat kérdése. Így ha nem rögzítjük azt, hogy két dolognak milyen tekintetben és mennyire kell hasonlóknak lennie ahhoz, hogy egy fajtaba tartozzanak, a hasonlósági viszonyok nem határozzák meg a fajtakat. A valóság maga pedig nem fogja ezt rögzíteni. Ebben Lakinak tökéletesen igaza van. Azt is elfogadom, hogy e hasonlósági kritériumok történetileg változnak. Csakhogy ezek a megfontolások nem mondanak ellent a realizmusnak, vagyis annak, hogy a szavak referenciáját a valóság ismeretlen vonásai is befolyásolják. Csak akkor mondanának ellent neki, ha realista azt állítaná, hogy a nyelvhasználati szokásainkban kanonizált hasonlósági kritériumok nem számítanak, a referencia *egyedül* a valóság szerkezetén múlik, vagyis csak akkor, ha elfogadná (2)-t.

Erre Laki azt felelhetné, hogy a realista, amennyiben nem fogadja el (2)-t, nem zárhatja ki a referencia megváltozását, s ezzel végső soron annak beismerésére kényszerül, hogy ő is csak minimálrealista. A realizmus önmagában valóban nem zárja ki a referencia megváltozását, de mégis van egy fontos különbség. A Kuhn- és Laki-féle felfogásban a valóság maga nem játszik szerepet a referencia meghatározásában, egyedül mentális moduljaink és nyelvhasználati szokásaink számítanak – a valóság legfeljebb figyelmeztet arra, ha ezekkel valami nincs rendben. Ebből következően, a mentális modulok és nyelvhasználati szokások radikális változásai automatikusan referenciaváltozással járnak, hiszen a referenciát egyedül ezek határozzák meg. A realizmus szerint azonban a valóság igenis beleszól a referenciába, ezért a mentális modulok változása nem eredményez automatikusan referenciaváltozást. Innentől kezdve pedig esetről esetre kell eldönteni, hogy megváltozott-e a referencia.

Lássunk egy példát. Laki szerint a „bolygó” szó referenciája a kopernikusi világkép elfogadásával megváltozott (141). A szó korábbi használata előfel-

tételezte a geocentrikus világgépet, s a Föld maga nem is számított bolygónak. A kopernikuszi világgép győzelmével a csillagászatban számtalan dolog megváltozott, egyebek között azok a kritériumok is, amelyek alapján egy égitest bolygónak minősült. Ha a referencia egyedül a kritériumokon múlik, akkor egy efféle radikális változás megváltoztatja a referenciát. A realista ellenben úgy gondolja, hogy a szavak referenciájának meghatározásában a valóság is közrejátszik. Azt kérdezi, hogy léteznek-e a valóságban olyan, egymásra bizonyos vonatkozásokban hasonlító égitestek, melyeket eleink „bolygónak” neveztek. Igen: a Merkúrt, a Vénuszt, a Marsot stb. eleink is „bolygónak” nevezték. Való igaz, azt az állítást, hogy „A Föld bolygó” hamisnak, sőt egyenesen groteszknek tartották, de ez nem befolyásolja azt, hogy a „bolygó” szóval mire referáltak. Téves csillagászati elméletet vallottak, téves kritériumokat használtak a bolygók azonosítására, ezért nem csoda, hogy tévedtek, amikor úgy gondolták, a Föld nem bolygó. Ez nem valamiféle „buta” tévedés volt, hiszen egy olyan felfogáson alapult, amely komoly intellektuális teljesítmény volt. De akkor is tévedés volt. E példával nem azt akarom mondani, hogy a referencia nem is változhat meg. Vannak olyan esetek, amikor sokkal bonyolultabb elmondani, hogy eleink egy adott szóval mire referáltak, ti. jelenlegi szókészletünkben nincs olyan szó, amely pontosan arra referálna, s vannak olyan esetek is, amikor azt kell mondanunk, hogy eleink nem referáltak, mert hogy az a dolog, amelyre referálni szándékoztak, nem létezik.

A példa ilyen elemzésével Laki két okból is elégedetlen lehet. Elsőként kifogásolhatja azt, hogy a realizmus szerint tévedhetünk szavaink referenciáját illetően, s nem is csak olyan módon, hogy valami zavaró körülmény vagy véletlen hiba folytán nem úgy használjuk a szavakat, ahogy azt nyelvi normáink szerint kellene, hanem tudatosan és elkötelezetten is tévedhetünk, ahogy eleink tévedtek, amikor a Földet kizárták a „bolygó” szó referenciájából. Ez a megállapítás teljes mértékben helytálló. Ha szavaink referenciája az általunk ismeretlen valóságon is múlik, ahogy a realista állítja, akkor tévedhetünk szavaink referenciájával kapcsolatban. A realista azonban nem talál ebben semmi kivétnevelőt. Szavaink a valóságra vonatkoznak, s a valóságtól nem várhatjuk el, hogy tökéletesen illeszkedjék nyelvhasználati habitusainkhoz. Annak, hogy a valóságra referáljunk, az az ára, hogy szemantikai ítéleteink, csakúgy, mint egyéb ítéleteink, hamisak lehetnek. Kockázat nélkül nincs igazság és nincs referencia.

Másodszor Laki azzal az ellenvetéssel élhetne, hogy a példa előfeltételez egy teljesen hamis képet. Azt, hogy közvetlenül hozzáférünk a valósághoz, hogy mintegy Isten nézőpontjából szemlélődhetünk, mintha egyfelől adott volna a nyelv, másfelől maga a valóság, s mi közvetlenül látnánk, hogy a kettő között mi a kapcsolat. De hát nem ez a helyzet: mi is be vagyunk zárva nyelvünkbe, s nem tudjuk szavaink referenciáját másként azonosítani, mint az erre szolgáló kritériumaink felhasználásával. Az utóbbi megállapítás helytálló. A valóságról csak annyit tudunk, amennyit legjobb elméleteink mondanak. A példa azonban

nem feltételezi a közvetlen hozzáférést. Adott a valóság a mi nézőpontunkból, a legjobb nézőpontból, amit ismerünk. Amikor azt vizsgáljuk, hogy eleink szavai mire referáltak, szükségképpen azt vizsgáljuk, szavaik hogyan kapcsolódtak az általunk ismert valósághoz. Miért az általunk ismert valósághoz, s miért nem ahhoz a valósághoz, melyet eleink ismertek? Azért, mert a mi elméleteink jobbak. Saját elméleteink mentesek azoktól az empirikus problémáktól, melyek végül a korábbi elméletek bukásához vezettek.

Ezek után már érthető, hogy az imént körvonalazott realizmus miért nem minimálrealizmus. Nem azért, mert másként vélekedik a referenciaváltozás lehetőségéről. Való igaz: ha realista szemmel tekintjük végig a tudomány történetét, abban kevesebb referenciaváltozást fogunk látni, mint ha kuhniánus szemmel tesszük ugyanezt. De a realizmust valójában az különbözteti meg a minimálrealizmustól, hogy úgy tartja: amennyiben sikerül referálnunk, akkor valóságos dolgokra referálunk. „Valóságon” itt pedig azt értem, amit Laki „realitásnak”, nem pedig „világnak” nevez, azt, ami tudatunktól függetlenül létezik, nem pedig azt, amit mentális moduljaink az érzéki ingerekből legyártanak. A valóság tehát nem csupán „általános empirikus nyomásként” vagy „egy bizonyos keménységként” lép be, jelezvén, hogy fogalmi sémánkkal gondok vannak. Ennél sokkal erőteljesebb szerepe van: ez szolgáltatja azokat az entitásokat, melyekre szavaink referálhatnak. (Ami, hadd ismételjem magam, nem azt jelenti, hogy a valóság nyelvhasználati szokásainktól függetlenül, önmagában is képes megszabni a szavak referenciáját.)

S most rátérek a harmadik kritikai megjegyzésre. Eddig amellet érveltem, hogy Laki tévesen feltételezi a realistáról, hogy elfogadja (1)-et és (2)-t, s hogy érvei hatástalanok egy olyan realista ellen, aki nem fogadja el (2)-t. De mi a helyzet (1)-gyel, a metafizikai realizmussal? A metafizikai realizmus Putnam Laki által idézett megfogalmazásában azt állítja, hogy „a világ elmefüggetlen objektumok meghatározott összességéből áll”, s „annak, hogy »milyen a világ« pontosan egy igaz és teljes leírása van” (114). Laki erre a felfogásra gondol, amikor az igazság korrespondenciaelméletet bírálja.<sup>3</sup> E felfogás szerint adott a valóság, s annak csak egyféleképpen lehet megfelelni. Ha nézeteink ilyen módon felelnek meg a valóságnak, igazak, ha nem, hamisak. Ez a kép feltételezi, hogy a megismerési tevékenység kétpólusú. Van a valóság egyfelől, s a vannak a kijelentések, hitek, elméletek stb. másfelől. Laki ezzel szemben úgy gondolja, van egy harmadik pólus is, jelesül, hogy mire kell nekünk a tudás. Némileg más összefüggésben ezzel a biológiai példával él: a légynek olyanféle ismeretekre van szüksége, amelyekkel egy légy életben maradhat és szaporodhat, és a denevér ultrahangos észlelő apparátusa által szolgáltatott ismeretekkel semmire sem menne (86–87). Ugyan-

<sup>3</sup> A Laki által idézett passzus így folytatódik: „Az igazság valamiféle korrespondencia referenciát foglal magába az egyik oldalon a szavak és gondolat-jelek, a másikon a külső dolgok és ezek halmazai között”. Hilary Putnam, id. mű, 49.

így, a tudományos megismerés során sem egyszerűen tudásra akarunk szert tenni, hanem olyan tudásra, amely eleget tesz bizonyos gyakorlati követelményeknek. E gyakorlati követelmények pedig különböző történeti szituációkban különbözők lehetnek. Mivel akár a valóság egyazon szeletével kapcsolatban is különböző gyakorlati céljaink lehetnek, nem csupán egyetlen megfelelési mód van, s ezért nem csupán egyféleképpen ítélnünk nézeteink igaz vagy hamis voltáról. Így a metafizikai realizmussal és az annak megfelelően értelmezett korrespondenciaelmélettel járó kizárólagossági igény jogtalan. Vagyis Laki pragmatista, s ebből fakadóan pluralista álláspontot képvisel az igazsággal kapcsolatban.

Ez az álláspont számomra is vonzó. Úgy gondolom azonban, hogy a tudomány fejlődése alapján nem lehet emellett, és a metafizikai realizmus ellenében érvelni. Laki érve, nagyon sarkosan megfogalmazva az, hogy a tudomány fejlődése során a változásokat olykor pragmatikus tényezők motiválják, s e pragmatikus tényezők szerepéről a metafizikai realizmus, amely kétpólusúnak tekinti a megismerést, nem tud számot adni. Például a kémiában használatos klasszifikációs kritériumok megváltoztatásáról, amelyre korábban már utaltam, azt írja, hogy „nem empirikus kérdés volt, elsődlegesen nem tudományos felfedezésektől függött, hanem a tudósok döntésétől” (138). De mi is itt a pragmatikus szempont?

Azáltal, hogy módosították az erre vonatkozó konvenciót, egyszerűbb, egységesebb leírást tudtak adni a megfigyelhető jelenségekről, s ez a gyakorlati tény motíválta döntésüket (uo.).

Ahogy a normál szakaszban működő kategória-rendszert azért cserélik le a forradalmak idején, mert felbukkannak olyan tapasztalatok, amelyek asszimilálására a régi fajtastruktúra nem alkalmas, ugyanúgy az új fajtaszerkezet kialakításánál is gyakorlati szempontok játszanak szerepet (146).

A cél nem a realitás új, a tényleges elrendeződésnek megfelelő leírása, hanem kifejezetten a konkrét, a régi paradigmában megoldhatatlannak bizonyult problémák megoldása (147).

A sikeresség nem implicál korrespondenciaként fölfogott igazságot vagy konvergens realizmust [...]. Azt föltételezi, hogy képesek vagyunk úgy rendelni relevanciaértékeket a realitás megfigyelhető aspektusaihoz, hogy az ily módon konstituált fajtaszerkezet fontos megfigyelési és kísérleti adatok koherens és konzisztens elméletbe rendezését tegye lehetővé. Az ebben az értelemben vett sikeresség feltétele, hogy az elmélet a realitás általános nyomásának megfelelően alakuljon, de mivel kevésbé specifikált, e nyomásnak természetesen sokféleképpen lehet megfelelni. Minden elmélet csak bizonyos aspektusoknak felel meg, azoknak, melyeket a tudományos közösség relevánsnak ítél (150–151).

Ha jól értem, a gyakorlati szempontot mindig azok a konkrét jelenségek vagy problémák jelentik, amelyeknek megértésére a kutatók az adott történeti szituációban törekednek. Ha van két kutató, s az egyik  $a$ ,  $b$  és  $c$  problémát kívánja megoldani, míg a másik  $A$ ,  $B$  és  $C$  problémát, akkor ki-ki a saját problémáin méri le, hogy egy adott elmélet megfelel-e a valóságnak. Azt az elméletet, amely a nagybetűs problémákat megoldja, a kisbetűs problémákkal küszködő kutató teljesen haszontalannak találhatja, és megfordítva.

Először is tegyük félre az olyan eseteket, amikor a problémák azért különböznek, mert a valóság más-más szeleteire vonatkoznak, mint például a biológiai és a meteorológiai problémák. Talán az ilyen esetek alapján lehet érvelni a metafizikai realizmus ellen, mondván, hogy a különböző problémamegoldások nem integrálhatók egyetlen elmélet keretében, de Laki nem ilyen alapon érvel. Számára azok az esetek az érdekesek, amikor a kutatók úgy értelmezik egymást, hogy a valóság egyazon szeletét vizsgálják, s megoldási kísérleteiket rivális, egymást kizáró elgondolásoknak tartják, de problémáik mégis különböznek. A ptolemaioszi csillagász problémája az, hogy olyan leírást adjon a bolygók mozgásáról, amelyben a bolygómozgások egyenes körmozgások eredői, a modern csillagász viszont olyan leírást kíván adni, amely összhangban van a gravitációs elmélettel és a relativitáselmélettel. Vagy: az arisztotelaiánus fizikus azt vizsgálja, hogyan mozog az a tárgy, melyet hozzákötünk valamihez, s ily módon megakadályozzuk abban, hogy eljusson természetes helyére, míg a galileiánus fizikus az ingamozgást kívánja megérteni. Csakhogy ilyenkor a problémák különbözőségének hátterében az elméletek különbözősége áll. Maga a probléma csak akkor fogalmazható meg egy adott módon, ha elfogadjuk egy bizonyos elméletet. A metafizikai realista pedig pontosan erre a tényre hivatkozhat. Azt mondhatja, hogy a problémafelvetések maguk is helyesek, illetve helytelenek lehetnek aszerint, hogy az elmélet, amelyet feltételeznek, igaz vagy hamis. A problémák pluralitásából ezért nem azt a következtetést szűri le, hogy a valóságnak többféle módon is meg lehet felelni. Az az elmélet ugyanis, amely a rosszul feltett problémára kínál megoldást, nem más módon felel meg a valóságnak, mint riválisa, hanem egyáltalán nem felel meg neki. Némileg absztraktabban fogalmazva, a nehézség az, hogy a problémák nem képeznek valamilyen harmadik pólust, amely meghatározná, hogy pontosan milyen megfelelést szeretnénk elérni elméleteink és a valóság között, hanem az elméleti pólushoz tartoznak.

Erre Laki felelhetne valami olyasmit, amit maga Kuhn is mond. Jelesül, hogy az olyan vitákban, amelyek nem egyszerűen a válaszok, hanem a helyes problémafelvetések körül is forognak, a tudósok nem pusztán elméletek között választanak. Ilyen esetekben aközött is választaniuk kell, hogy milyen módon műveljék a tudományt, azaz hogy milyen fogalmi és kísérleti eszközökkel, milyen

problémákat, milyen szabályok szerint kívánnak megoldani. Azt kell eldönteniük, hogy milyen tudományos gyakorlatot folytassanak. Erre azonban a metafizikai realistának kész válasza van: a gyakorlatok nem egyenértékűek, s helyességük azon múlik, hogy melyik feltételez, illetve produkál a korrespondencia értelmében vett igaz állításokat. Amit Kuhn és Laki gyakorlati szempontnak nevez, amelyre nem alkalmazható a korrespondenciaelmélet, az a metafizikai realizmus szemében elméletet involvál, s nem vonható ki a korrespondenciaelmélet hatálya alól. Tévedés ne essék: ezzel nem azt akarom mondani, hogy a metafizikai realizmus helyes, hanem csak azt, hogy a tudományfejlődés tényei alapján nem lehet fogást találni rajta.

## A hermeneutika helye Thomas Kuhn gondolkodásában

### Megjegyzések Laki János Kuhn-értelmezéséhez

Laki János arra tesz kísérletet, hogy monografikus alaposággal rekonstruálja Kuhn tudományfilozófiai nézeteit, illetve beillesse a kuhni fordulatot a tudományról való gondolkodás 20. századi történetébe. A könyv jól sikerült vállalkozás: Laki felkészülten ábrázolja azt a filozófiai, tudományelméleti szituációt, melybe Kuhn nézetei berobbantak, markáns és eredeti képet fest a kuhni gondolkodás fő elemeiről, belső összefüggéseiről. Különösen elgondolkodtató és izgalmas, ahogy Laki a tudományos forradalmak természetével, a tudomány fejlődésével, illetve az intézményi racionalitás szerepével kapcsolatos kuhni nézeteket rekonstruálja.

Írásomnak nem az a célja, hogy Laki könyvét egészében értékeljem, hanem csak arra szeretnék vállalkozni, hogy egy számomra különösen fontos aspektusból, Thomas Kuhn és a hermeneutika viszonyának, a kuhni tudományfilozófia hermeneutikai aspektusainak a kérdése szempontjából vizsgáljam az általa leírtakat. Elsődlegesen azt szeretném vizsgálni és vitatni, ahogy Laki e problémakört a könyvében konkrétan tárgyalja, de aztán arra az általános szemléletmódbeli kérdésre is ki szeretnék térni, amely arra vonatkozik, hogy vajon milyen filozófiai, illetve tudományos háttér segíthet leginkább a kuhni gondolatok rekonstrukciója számára.

Laki a hermeneutika szerepét a kuhni gondolatvilágon belül szigorúan lehatárolva vizsgálja a könyvében. E lehatárolás már a könyv felépítésében is megnyilvánul, hiszen e kérdéskört kizárólag a tudománytörténet problematikájával foglalkozó rész egyik – „Hermeneutika” címet viselő – alfejezetében tárgyalja. Laki tartalmi vonatkozásban is egyértelmű: elemzésében kizárólag a „kuhni tudománytörténész” (44) gyakorlatával foglalkozik, vizsgálatának magától értetődő feltételezése, hogy Kuhn kapcsán *csakis egy vonatkozásban, a (tudomány)történeti megértés keretei között vetődik föl e hermeneutika problémája.*

Laki rekonstrukciója szerint Kuhn a hetvenes évek közepétől kísérletet tett arra, hogy a hermeneutikát mint módszert hívja segítségül az inkommenzurábilis gondolatrendszerek között közvetítő tudománytörténészek számára. E vázlatosan kidolgozott tudománytörténeti hermeneutika kiindulópontja az úgynevezett



jóindulat elve, amely Kuhnnál annyit jelent, hogy egy számunkra már idegenné vált tudománytörténeti szöveghez úgy közelítünk, hogy annak – sajátunktól eltérő – racionalitását eleve feltételezzük. Ebből az elvből már következik, hogy a szövegértés nehézségeit, anomáliáit nem a szerző gyengeségének fogjuk tekinteni, hanem az erőfeszítéseinket éppen e pontok megértésére irányítjuk – abból a célból, hogy a hagyományos szöveg valódi értelmét felderíthessük. A szövegértés során fellépő anomáliák kiküszöbölésével közelíthetünk a valódi célunkhoz, az áthagyományozott szöveg eredeti értelmének konzisztens összefüggésként való rekonstruálásához. E hermeneutika pozitivistikus jellegű, amennyiben a tudománytörténeti tényeknek Kuhn túlzó fontosságot tulajdonít.<sup>1</sup>

E tudománytörténeti hermeneutika jelentőségét Laki több szempontból is megkérdőjelezi. Egyfelől a kuhni hermeneutikát – Hackingra támaszkodva – szembeállítja Gadamer hermeneutikájával, mondván, hogy Kuhnnak nem tett szert volna Gadamer hermeneutikája – már ha ismerte volna –, másfelől pedig Kuhn azzal, hogy „tétova és szisztematikusan végig nem vitt” (45) módon vonta be a hermeneutikát, a lehetséges interpretációk sokféleségét és bizonytalanságát vitte be a tudománytörténeti értelmezésbe.

A következőkben néhány ponton vitába szeretnék szállni Laki rekonstrukciójával és értékelésével. A legfontosabb kérdésnek azt tartom, hogy milyen hermeneutikát tulajdonítunk Kuhnnak, vajon célszerű-e valamely részproblematikára lokalizálnunk a hermeneutikai gondolkodásmód megjelenését, avagy a kuhni gondolkodásmód egészét hermeneutikai jellegűnek tartjuk. További kérdés, hogy melyik hermeneutikához nyúlunk vissza a filozófiai tradíción belül. Gadamer – a szellemtudományi önmegértéshez visszaforduló – hermeneutikáját tekintjük-e referenciapontnak, vagy inkább a – jóval radikálisabb és tudományfilozófiai szempontból relevánsabbnak tűnő – heideggeri hermeneutikához fordulunk?

*Laki tehát két módon is lehatárolja Kuhn és a hermeneutika érintkezésének problematikáját:* Gadamerre szorítkozik, továbbá kizárólag a tudománytörténeti megértés hermeneutikai vonatkozásait vizsgálja – a magam részéről mindkét behatárolást problematikusnak érzem. Továbbá úgy vélem, hogy – részben e lehatárolás következtében – még azok a mozzanatok is vitathatók Laki értelmezésében, ahol Laki – a kuhni megjegyzések nyomán – kapcsolatot lát Kuhn és a hermeneutika között.

Közbevetőleg érdemes megjegyezni, hogy Kuhnnak a hermeneutikához fűződő viszonya valóban sajátos természetű, hiszen alapvető gondolatait úgy fejtette ki, hogy nem ismerte a hermeneutikát – mint 1977-ben írja: „a »hermeneutikai« kifejezés [...] öt évvel ezelőtt még nem tartozott a szókincsemhez” (Kuhn 1988. 36), emiatt a hermeneutika és Kuhn viszonyában nem azt tekinteném elsődle-

<sup>1</sup> „A hermeneutikai értelmezés e pozitivista jellegének feltárása mintegy visszamenőleg világgossá teszi, hogy mi is volt a baj a »történeti tény« kuhni felfogásával” (44).

ges kérdésnek, hogy ő sok esetben *utólagos, önértelmező módon*<sup>2</sup> miként azonosítja a „hermeneutikait” saját gondolatvilágában. Meg kell tehát különböztetnünk Kuhn utólagos önértelmezésének problémáját attól a sokkal fontosabb kérdéstől, hogy az a tudományértelmezés, amelyet Kuhn – a hermeneutikai tradíció ismerete nélkül – kidolgozott, *egészében hermeneutika-e, s ha igen, akkor ennek milyen filozófiai következményei vannak.*

A legélesebben talán az alkalmazás hermeneutikai elvével kapcsolatos értelmezésben érhető tetten feszültség Kuhn tényleges gyakorlata és Laki interpretációja között. Laki Kuhn tudománytörténeti hermeneutikájának módszertanát elemezve arra a következtetésre jut, hogy e módszertan „kizárja a kuhni hermeneutikából az alkalmazás problémáját” (43). Véleményem szerint azonban az applikáció kérdése több szempontból is kitüntetetten fontos Kuhn és a hermeneutika kapcsolatának vonatkozásában, akkor, ha a tudománytörténet kuhni megközelítését vizsgáljuk, mind pedig abban az esetben, ha Kuhn tudományfilozófiájának egészét tekintjük.

Jól ismert, hogy a kuhni tudományfilozófiai fordulat egészének szempontjából kulcsfontosságú mozzanat a paradigma-fogalom bevezetése, amelynek egyik alapjelentése éppen az, hogy olyan példázatokra, konkrét feladatmegoldásokra vonatkozik, „melyeket a tudományos közösség a tudományos munkában min-tának tekint” (21). Laki teljes joggal nyomatékosítja – *a hermeneutika kérdésétől azonban teljesen elválasztva* – a paradigma fogalmának e jelentését, ki is kiemeli, hogy Kuhn a paradigma ezen értelmét „tekinti a könyv legújszerűbb és legkevésebbé megértett gondolatának” (uo). A példázatok, konkrét feladatmegoldások azonban mi mások, mint alkalmazások? A paradigma-fogalom e meghatározásával Kuhn valójában azt állítja, hogy *az alkalmazás a tudományos elméletek elsajátításának, megértésének döntő mozzanata.* A hermeneutika éppen azt hangsúlyozza, hogy valaminek a megértése az alkalmazása során valósul meg. Amint Gadamer írja: „a megértés [...] mindig egyben alkalmazás” (Gadamer 1990. 314, magyarul 219). Ez a Heideggerre visszanyúló gondolat pontosan azt jelenti, hogy valaminek, például egy teóriának a megértése elvileg nem választható el az elmélet konkrét alkalmazásaitól, az alkalmazások kitüntetett szerepet játszhatnak egy elmélet megértésében, elsajátításában.

Amikor Laki a kuhni hermeneutikától elvitatja az alkalmazás mozzanatát, akkor azt természetesen nem a paradigma, hanem a történeti megértés vonatkozásában teszi. Csakhogy ha megfontoljuk, hogy milyen kapcsolat van a tudomány-

<sup>2</sup> Kuhn az imént már idézett 1977-es írásában – tanulmányainak német nyelvű kiadásához írt előszavában – is 1947-es Arisztotelész-értelmezésének nehézségeire tekint vissza, amikor szóba hozza a hermeneutika problémáját. Ugyanakkor fontos hangsúlyoznunk, hogy hermeneutikai belátásainak érvényét kifejezetten *tudományképeinek egészére* vonatkoztatja: „Esetemben azonban a hermeneutika felfedezése nem csak ahhoz vezetett, hogy a történelem mint valami figyelemre méltó tárgy jelent meg a számomra. A hermeneutika legközvetlenebb és legdöntőbb hatása inkább a tudományról kialakított képemet érintette” (uo. 34. sk.).

történeti példák és a kuhni tudományfilozófiai elmélet között, akkor éppen az alkalmazás kérdésébe ütközünk. Értelmezésem szerint amikor Kuhn tudománytörténeti példákkal operál, akkor tulajdonképpen *alkalmazott tudományfilozófiát* művel. Tudománytörténeti szempontból sok esetben talán vitathatók ezek a példák, ám látnunk kell a funkciójukat: valójában tudományfilozófiai alkalmazásokról van szó. Kopernikusz vagy Lavoisier története eszerint nem más, mint a paradigmaváltás példázata, Kuhn elméletének konkrét alkalmazása, amelynek tanulmányozása révén megérthetjük a kuhni tudományfilozófiát – e példák, alkalmazások pontosan azt a szerepet játszhatják, melyet más tudományokban is a példázatok játszanak: tanulmányozásuk révén a tudósok, jelen esetben a tudományfilozófusok „kezdik másként látni a tudományukat” (Kuhn 1984. 117).

Gadamer és Kuhn szembeállításához az is hozzájárulhat, hogy Laki a történeti megértés gadameri elméletének bemutatása során a hangsúlyt eltolja az értelmező szituáltságának mozzanatára; pedig Gadamer elmélete kiegyensúlyozott e tekintetben, s nem csak azt hangsúlyozza, hogy a hermeneutika feladatai közé tartozik az értelmező hermeneutikai szituációjának értelmezése (vö. Gadamer 190. 310. skk, magyarul 216. sk), hanem a „történeti hagyomány [...] *maga is megszólal a maga igazságában*” (Gadamer 1990. 3, magyarul 22)<sup>3</sup>. Gadamer egyoldalú értelmezéséhez az is hozzájárulhatott, hogy Laki egy helyen kevésbé sikerült fordításra támaszkodott; a vitatott – Bonyhai Gábor által fordított – mondat a következő: „a történeti megértés voltaképpen tárgyát nem események képezik, hanem azok »jelentősége«”. A német eredetiben a Bedeutung kifejezés szerepel (Gadamer 1990. 334, magyarul 231), melyet szótári szempontból lehet ugyan jelentőségnek fordítani, de ez mégis félrefordításnak tekinthető: a szövegösszefüggésből és az egész gadameri elméletből egyértelműen kiviláglik, hogy jelentésről van szó – vagyis a történeti megértés az illető szöveg jelentését próbálja meg felderíteni. A gadameri gondolat értelmezése, miszerint pusztán arról lenne szó, „hogy milyen jelentőséget tulajdonítunk a tudománytörténet korábbi szerzőinek” (44), félreértésnek tekinthető.<sup>4</sup>

Van azonban egy fontos pont, melyben egyetértek Laki értelmezésével. Számomra is nyilvánvalónak tűnik, hogy Kuhn *saját pozíciójának értelmezésekor* nem alkalmazott következetesen hermeneutikai elveket, nem vette eléggé figyelembe saját pozíciójának a történetiségét, s könnyen lehet, hogy pozitivista előítéleteket táplált a tudománytörténeti tényekkel kapcsolatban. Kuhn önreflexív elgondolásaitól tényleg nem áll távol az, ahogy Laki jellemzi a kuhni „hermeneutikát”: „célja annak feltárása, hogyan gondolkodtak »ők«, s nem annak leírása, mit mond számunkra a szöveg ma” (43). Ugyanakkor azt is világosan látnunk

<sup>3</sup> „Ezért a hermeneutikai hozzáálláshoz szükségképp hozzátartozik a történeti horizont felvázolása, mely különbözik a jelen horizontjától” (Gadamer 1990. 311, magyarul 217).

<sup>4</sup> Gadamer néhány sorral a vitatott mondat előtt például a következőképpen fogalmaz: „Nem érvényes-e valójában minden szövegre, hogy azt kell megérteni, amit mond?” (Gadamer 1990. 334, magyarul: 231).

kell, hogy Kuhn tényleges gyakorlata viszont pontosan az volt, hogy bemutassa, mit mond ma – tudományfilozófiai szempontból – számunkra a szöveg.

Kuhn paradigma-fogalmának hermeneutikai aspektusait érintő megfontolásaink révén már túlléptük azt a határt, amelyet Laki szabott, amennyiben Kuhn és a hermeneutika kapcsolatát kizárólag a tudománytörténeti megértés vonatkozásában vizsgálta. Laki lehatárolásával szemben véleményem szerint Kuhn tudományfilozófiájának egésze hermeneutikai jellegű. Természetesen a kuhni tudományfilozófiának a hermeneutika szempontjából történő értelmezése az egyik lehetséges interpretációnak tekinthető, amelynek gyengéi és erősségei vannak, ám mindenképpen az *egyik* lehetséges Kuhn-értelmezés. Kuhnnak nincs olyan jelentős gondolata, melynek ne lenne a hermeneutika által megfogalmazható verziója, s van jó néhány olyan kérdés is, ahol a hermeneutikai értelmezés sokkal plauzibilisebb megoldással tud előállni, mint az egyéb Kuhn-értelmezések.

E vonatkozásban Laki könyvét úgy tekinthetjük, mint amely egy – legalábbis számomra – *fontosnak tűnő vitát elmulasztott*, ami különösen azért hiányzik, mert Laki írása kifejezetten jó: értelmezése sok kérdésben eredeti és izgalmas megoldásokat tartalmaz. E rövid írásnak persze csak az lehet a célja, hogy jelezze a vita „helyét”, azokat a pontokat, ahol a hermeneutikai értelmezés valódi alternatívát kínál, illetve ahol Laki értelmezése természetes szövetségesének tekinthető. Gondolatmenetem zárásaként mindkét esetre hoznék egy-egy példát.

A hermeneutika kitűnő szövetségese lehetett volna Lakinak a realizmussal kapcsolatos interpretációjában. Laki, miután megállapítja, hogy „jóformán teljes a konszenzus abban, hogy (Kuhn) nézetei *antirealisták*” (83), felhívja a figyelmet arra, hogy „valahogy mégis számot kellene vetni azzal a zavaró tényezővel, hogy Kuhn következetesen elutasítja az »antirealista« címkét” (84). Laki válasza a dilemmára úgy szól, hogy Kuhn „minimálrealista álláspontot képvisel, melyben a tudományfilozófust kantianus felfogása támogatja. Laki rekonstrukciójában megkülönbözteti a realitást a paradigmák által konstituált világoktól, melyek nélkülözhetetlenek ahhoz, hogy a realitás valamely metszetét megismerhessük: „a különböző paradigmák a realitás más-más aspektusát engedik megmutatkozni” (114). Ez a felfogás gyakorlatilag megegyezik azzal, amit Heidegger a húszas évek közepén dolgozott ki: a világ olyan értelem-összefüggés, amelynek a segítségével a létezőt<sup>5</sup> valamilyen módon felfedjük.<sup>6</sup> Persze hozzá kell tennünk, hogy Heidegger az ontológiai differencia gondolatával jóval mélyebbre hatolt,<sup>7</sup> mint Kuhn, de a konstrukció strukturálisan mégis párhuzamos: mindkét esetben a ha-

<sup>5</sup> Laki a realitás kifejezést használja, Kuhnnál pedig a két értelemben felmerülő világ-fogalom egyike jelenti ezt: „bár a paradigma megváltozásával a világ nem változik meg, a tudós azután egy másik világban dolgozik” (Kuhn 1984. 165).

<sup>6</sup> Ezt az értelmezést részletesen a következő helyeken fejtettem ki: Schwendtner 2000a. 93. skk, Schwendtner 2000b. 141. skk.

<sup>7</sup> E fogalomhoz lásd Schwendtner 2001. 79. sk.

gyomány áltai konstituált világ és a realitás egy szelete korrelációs viszonyának (fenomenológiai)\* alapelvéről van szó.

A hermeneutikai, fenomenológiai szemlélet alkalmazásának egyik legnagyobb előnye véleményem szerint abban áll, hogy filozófiailag, módszertanilag egységes keretek között tudja vizsgálni azokat a problémákat, melyeket egyébként heterogén módszertani háttérrel, más-más tudományoktól kölcsönzött nyelven tud maga Kuhn, illetve tudnak Kuhn értelmezői kifejezni. Egy példa erre: Laki értelmezése szerint „a fejlődést mozgató »lényegi feszültség« ennek megfelelően *nem individuális kutatók pszichológiai állapota*, hanem a tudományt művelő individuumokból összeálló *közösség természetes sokféleségének megnyilvánulása*” (208). Eszerint Laki a tudományban rejlő lényegi feszültség problémáját a pszichológia és a szociológia diszciplináris szembeállításával értelmezi, s a kérdést úgy dönti el, hogy a közösségi (szociológiai) értelmezés mellett foglal állást. Csakhogy a Laki által idézett szövegrész mindkét értelmezési lehetőséget magában foglalja, nem csak a Laki által idézett közösségi verziót erősíti meg,<sup>9</sup> hanem Kuhn a főszövegben azt is írta, hogy „az eredményes kutatónak egyidejűleg kell a tradicionalista és a képromboló tulajdonságait felmutatnia” (Kuhn 1977. 227), melyet csak kiegészített és pontosított a lent idézett lábjegyzetével.

Azt gondolom, hogy kár lenne az individuális/egzisztenciális, illetve a közösségi mozzanatokot *diszciplinárisan* szétválasztani, mondván az egyik pszichológiai, a másik ellenben szociológiai jellegű, hanem ezeket a mozzanatokot *módszertanilag egységesen, összefüggésükben* érdemes tárgyalni. Ha a hermeneutika nyelvén fogalmaznánk meg ezeket a dilemmákat, rengeteg felesleges vitától és szembeállítástól kímélnénk meg Kuhn értelmezését. Kuhn persze különféle diszciplináktól vett ötleteket, metaforákat, de ha elfogadjuk azt a plauzibilis feltevélezt, hogy filozófiát csinált, akkor véleményem szerint *nem érdemes* ezeket az eredetileg pszichológiai, szociológiai, tudománytörténeti eszméket, képeket a származási helyükhöz, a *nem filozófiai tudományokhoz visszakötni*. Ha egyszer e gondolatok *filozófiai gondolatmenetek* részeivé válnak, akkor célszerű lenne egységesen, filozófiai szempontok alapján értékelni, megítélni őket. Véleményem szerint Kuhn értelmezésekor gyakran előfordul, és sok konfúziót okoz, hogy *hiányzik* az egységes filozófiai nyelv és látásmód, ezért a különféle szaktudományok mentén szétcincálják és szembeállítják a kuhni gondolatokat, pedig *egységes filozófiai tárgyalás esetén a felmerülő nehézségek egy része magától megoldódna*.

\* Véleményem szerint pontatlan kantianusnak nevezni ezt az ízig-vérig fenomenológiai koncepciót.

<sup>9</sup> A Laki által idézett lábjegyzetszöveg a következőképpen hangzik: „inkább a szakmai csoport, mint az individuum az, melynek egyidejűleg rendelkeznie kell a két tulajdonsággal. A csoport egyes tagjai lehetnek inkább tradicionalisták, mások inkább újítók, s ennek megfelelően különbözőképpen járulnak hozzá a csoport teljesítményéhez. Mindazonáltal az oktatás, az intézményi normák és az elvégzendő feladat természete együttesen biztosítják, hogy többé vagy kevésbé mindkettő áthassa a csoport valamennyi tagját” (Kuhn 1977. 227. sk., Laki 2006. 208).

A Heidegger által kidolgozott hermeneutikai fenomenológia az egyik lehetőség arra, hogy ezt a filozófiai háttérrel nyújtsa, s a Kuhn-interpretációk állandó kihívása az a kérdés, hogy létezik-e másik olyan filozófiai elmélet, amely képes arra, hogy a kuhni elmélet *átfogó és egységes filozófiai értelmezését* nyújtsa.

#### IRODALOM

- Gadamer, Hans-Georg 1990. *Gesammelte Werke 1, Wahrheit und Methode*. Tübingen, Mohr. Magyarul: Hans-Georg Gadamer 1984. *Igazság és módszer*. Ford. Bonyhai Gábor. Budapest, Gondolat. 1984.
- Heidegger, Martin 1993. *Sein und Zeit*. Tübingen, Niemeyer. Magyarul: uő 2001. *Lét és idő*. Ford. Vajda Mihály, Angyalosi Gergely, Bacsó Béla, Kardos András, Orosz István. Osiris, Budapest.
- Kuhn, Thomas S. 1984. *A tudományos forradalmak szerkezete*. Ford. Bíró Dániel. Budapest, Gondolat.
- Kuhn, Thomas S. 1977. *The Essential Tension*. Chicago–London, The University of Chicago Press.
- Kuhn, Thomas S. 1988. *Die Entstehung des Neuen*. Hrsg. L. Krüger. Frankfurt am Main, Suhrkamp.
- Schwendtner Tibor 2000a. *Heidegger tudományfelfogása*. Budapest, Gond–Osiris.
- Schwendtner Tibor 2000b. Thomas Kuhn és a hermeneutika. *Replika* 41–42. 139–148.
- Schwendtner Tibor 2001. Heidegger Husserl-kritikája és a létkérdés. In uő (szerk.) *Metszéspontok. A hermeneutika és a fenomenológia határvidékén*. Budapest, L'Harmattan. 75–97.

## Hogyan konstruáljuk a szerzőt és a világot?\*

Ritka alkalom egy szerző életében, hogy egyszerre négy kiváló pályatárs szentel figyelmet könyvének. Ha e figyelem nem is csupán formális, hanem valóban részletekbe menő, az persze veszélyes, hisz ki-ki a maga szakterületéhez tartozó problémákra koncentrál, olyan kérdésekre, melyekhez szükségképp jobban ért, mint a szerző, akinek ezek csak az általa megírni kívánt egész résztémái. Megnyugtatóan hatott hát rám Pléh Csaba böles megjegyzése, miszerint „Kuhnnek [...] több olvasata jelent meg [...] Laki [...] világosan érzékelteti, hogy az értelmezés attól függ, hogy melyik részre helyezzük a hangsúlyt.” Valóban az volt a törekvésem, hogy a standard Kuhn-értelmezést korrigáljam, s ennek eszközéül a hozzáférhető szövegekörpusz hangsúlyainak áthelyezését választottam.

Pléh a kognitív pszichológus szemüvegével olvassa Kuhnt, mint olyan szerzőt, aki a maga idejében pszichológiailag jól megalapozott tudományfelfogást képviselt, de az általa és róla folytatott vitáknak máig megvan a fontos tanulságokat hordozó utóélete. „Mindig valamit valami előtt látunk” – írja más összefüggésben, s elismerem, Kuhn figurája jobban kivethető lenne az utóbbi húsz év pszichológiai fejleményeinek gazdagabb háttere előtt. Mivel a már megjelent könyvekhez és tanulmányokhoz képest nem volt lényeges, új mondanivalóm, e háttér megrajzolását ráhagytam az egyre bővülő szakirodalomra.

Pléh Kuhn-olvasata tudásszociológiailag is érzékeny. A szövegnek azért nincs egyetlen értelme, mert az olvasók mint „hús-vér lények” „beágyazottságuknak megfelelően” helyezik ide vagy oda a hangsúlyokat. 1970-ben fontosabb volt a forradalmi Kuhn, „a radikális változás és a szabadság szövetségese”, mára élesebb kontúrokkal rajzolódik ki a konzervatív, aki szerint gondolkodásunkat episztémikus sémák, kognitív mintázatok szervezik. A társadalomfilozófiai olvasat viszont mára jóformán teljesen elveszett, s vele érezhetően csökkent a Kuhn által nem szakmai körökben kiváltott izgalom is. Pléh úgy érzi, a mai Magyar-

\* Köszönettel tartozom az OTKA-nak a szükséges kutatómunkához nyújtott támogatásért (K 72598).

országban talán a szükségesnél több is a rá fordított figyelem. Másik recenzió, Zemplén Gábor épp ellenkezőleg, kifejezetten szerencsésnek látja a témaválasztást. Nem pusztán azért, amiért a világban a Kuhn iránti érdeklődés nem csökken, hanem épp a magyar szellemi élet szempontjából: a Kuhn által fölvetett problémák tudománytörténészek, szociológusok, filozófusok, nyelvészek és pszichológusok közös területére esnek, s így hidat képeznek a távolodó diszciplínák között.

## I. AZ INTERPRETÁCIÓ ANOMÁLIÁI

Bár teljesen egyetértek Zemplén általános attitűdjével, konkrét elemzéseit olvasva helyenként úgy érzem, nem ad elég esélyt a szöveg jelentésintenciójának kibontakozására. Ilyenek például az anomália fogalmi tisztázására vonatkozó passzusok. „Kuhn – és Laki is – rendre objektív tényként utal az anomáliára”, írja, holott „az, hogy valami az anomália státuszával rendelkezik, [...] társadalmi folyamatok eredménye”. Értelmezését idézettel támasztja alá, s tudjuk, az idézetek valahol véget érnek. Valóban azt is írtam, hogy „az anomália [...] a paradigma alkalmatlanságát jelzi”, de a gondolat így folytatódik: „Ennek ellenére, *ameddig a bizalom fennáll*, még a makacs és elismert anomália sem okoz [...] válságot” (188 – mostani kiemelés). Vagyis pontosan az történik, amit a recenzió hiányol: a megoldási kísérleteknek ellenálló probléma társas értékelése, amelynek feladata annak eldöntése, hogy továbbra is fenn kell-e tartani az eszközökbe vetett bizalmat. Két mondattal később ez explicit módon ki is van mondva: „A huzamosan ellenálló rejtvény anomáliává történő átminősítése alapvetően *nem episztémikus* [...] hanem *bizalmi kérdés*” (uo.). Ezt követően, elég részletesen elemzem az eszközökbe vetett bizalom fenntartásának, megerősödésének vagy elmúlásának társas folyamataiból, a *kockázatredukáló* és *sikermaximalizáló* tudósok hozzáállásának különbségeiből kibontakozó *rivalizációt* és az ezt lezáró *konszenzuális* döntési folyamatot.

Bár meglepett a szöveg szándékolt értelmétől erősen eltérő interpretáció, nem hoznám szóba, ha ennek következtében a kuhni „anomália” és vele a „paradigmaváltás” fogalma nem tűnne a ténylegesnél popperianusabbnak, s ezzel nem homályosulna el a könyv egyik alapvető mondandója. Azokra az alfejezetekre gondolok, amelyek pillanatszerű *esemény* helyett társadalmi *folyamatként* írják le a tudományos változást, s ezzel megmutatják annak racionalitását. Fájjalom, hogy nem sikerült elég éles fénybe állítani a *bizonyítás* és *meggyőzés* különbözőségét, s a *bizalom lebomlásának mikrofolyamatait* vagy a *kockázatmegosztást* elemző alfejezetek nem elég erőteljesek ahhoz, hogy láthatóvá tegyék a tudományfejlődés „társadalmilag konstruált” voltát. Természetesen, e konkrét polémia jelentéktelen, de ha bizonyos félmondatok túlságosan előtérbe, mások túlságosan háttérbe kerülnek, az könnyen vezet új szerző konstruálásához.



Az a benyomásom, Zemplén túl könnyen fogadja el azt a megoldást, hogy Kuhnnek volt (legalább) két korszaka: előbb ezt mondta, aztán azt, s a kettőt nem feltétlenül kell összeegyeztetni, még akkor sem, ha kénytelenek vagyunk elismerni, hogy a változtatás következményeként megerősödik a parttalan relativizmus, amelytől évtizedeken át igyekezett elhatárolódní. Azt, hogy a késői írásokban előtérbe kerül a speciáció, Zemplén fordulatként fogja föl, s ezért hibának látja, hogy én a „késői” Kuhnt a „korai” és „középső” Kuhn értelmezéséhez használom. Szerintem ebben a konkrét esetben, jobban teljesít az általam választott interpretációs elv: Kuhnnek volt egy eredeti elképzelése a tudomány működéséről, amelyet finomított és pontosított, de alapvetően soha nem változtatott meg. A hozzáférhető szövegek nem alapozzák meg a TFSZ<sup>1</sup> és a késői írások evolúciófelfogásának azon az alapon történő szembeállítását, hogy a „korai” Kuhn mutáció-szelekció modellt használt, a „késői” viszont speciációt, s ezek eltérő szemantikai és racionalitás-koncepciót implikálnak.

Nem érzem kellően alátámasztottnak azt a megállapítást, hogy Kuhn „először” a késői szövegekben „fejtette ki a speciációs [...] modellt”. A *speciáció fogalma* (noha maga a *terminus* nem) jelen van már a TFSZ-ben is, s a „hivatásos tudóst más területeken dolgozó kollégáitól elválasztó egyre szélesedő szakadékról” beszélve, Kuhn jelezte, hogy szerinte „túl kevés figyelmet szentelnek e szakadék és a tudományos haladás [*scientific advance*] benső mechanizmusai közötti lényegi kapcsolatnak”<sup>2</sup>. A TFSZ-ben világosan megfogalmazódik, hogy a speciális diszciplínák létrejötté a vizsgálati területek *szűkülését* és a más területekre specializált csoportokkal való *kommunikáció megnehezülését* eredményezi.<sup>3</sup> Ennek fényében aligha tekinthetjük a „késői” Kuhnra maradó drámai felismerésnek, hogy a kutatás instrumentális és tematikus specializálódásának velejárója a diverzifikálódás.

Maga a „késői” Kuhn is utal arra, hogy a speciáció gondolata jelen van a „korai” műben: a taxonómiai változás során történik valami, „amit az SSR csak mellékesen említett. A forradalom után rendszerint [...] több kognitív specializáció vagy kutatási terület lesz, mint előtte volt.”<sup>4</sup> Az egyik ilyen „mellékes említés” lehet a következő: „a tudomány bizonyosan egyre mélyebbre hatol [s egyre kiterjedtebbé is válik] e kiterjedés főként a speciális diszciplínák számának növekedésében mutatkozik meg...”<sup>5</sup> A korai írások figyelmes olvasóját ezek után nem éri váratlanul, hogy a késői tanulmányokban ugyancsak a tudomány mélységében és szélességében

<sup>1</sup> A *tudományos forradalmak szerkezete*, az idézeteknél az eredeti szöveg 1970-es kiadására hivatkozom: SSR.

<sup>2</sup> SSR 21. – kiemelés L. J.

<sup>3</sup> „revolution narrows the scope of the community’s professional concerns”; „attenuates its communication with other groups, both scientific and lay” – SSR 170.

<sup>4</sup> Kuhn 1990a. 97. – kiemelés L. J.

<sup>5</sup> „science surely grows in depth”; – „in breadth as well”; „that breadth is manifest mainly in the proliferation of scientific specialities” (SSR 170. – kiemelés L. J.)

gében való fejlődésének gondolata merül föl, s a fejlődés e két módja egyáltalán *nem egymást kizáró*, hanem éppenséggel *komplementer*.

Kuhn egy 1990-es, éppen a gondolkodásában a TFSZ megjelenése után történt *változásokat áttekintő* előadásában azt mondta, hogy a TFSZ végén fölbukkanó evolúciós analógia „ugyanazon vagy részben azonos természeti jelenségekre vonatkozó régebbi és újabb tudományos vélekedések közötti viszonyra” vonatkozott. „Most – mondta – javasolnék egy második, kevésbé észlelt párhuzamosságot [...] mely a *tudományok* szinkrón *keresztmetszetére*, s nem *egy tudomány* diakrón *hosszmetszetére* vonatkozik”<sup>6</sup>. A gondolat egyértelmű: a mutációs-szelekciós modell egy tudomány időbeli, míg a speciációs modell „a tudományok” egyidejű artikulálódásának leírására használatos. Talán nem szükséges tovább sorolni azokat a szöveghelyeket, amelyek egyértelművé teszik, hogy „a folyamatról, mely által a proliferáció és lexikonváltás végbemegy” Kuhn nézetei valóban „nagyon közel állnak a *Structure*-ben kifejtettekhez”.<sup>7</sup> Ezért inkább elismerésnek, semmint kritikának veszem Zemplén megállapítását, miszerint „a késői szövegeket is a korai Kuhn evolúciós felfogásának rekonstrukciójára használ[om]”.

A mélységi és szélességi fejlődés kapcsolatának feltárása után más fény vetül az optimalizáció és a racionalitás kuhni fogalmára is. Az ezekkel kapcsolatos véleménykülönbségünk a dawkinsi „mém-fogalom” értelmezésével kezdődik. Zemplén szerint „a mémelméletet [...] az olyan [...] »irracionális« folyamatok magyarázatára lehet jól használni, mint a divatjelenségek...”, de nem alkalmas a tudományfejlődés racionális magyarázatára. Nem világos, miért lenne a divat „irracionális” jelenség. Ha egyszerűen nem-racionálisnak minősítenénk, láthatóvá tenné a mém és a paradigma lényegi hasonlóságát: mindkettőnek központi eleme az *utánzás*. A paradigmaelmélet döntő újítása a megelőző metodologiz-mushoz képest, hogy explicit módszertani szabályok helyett minták vezetik a kutatást. Ez valóban különbözik a szabály- és adatvezérelt racionalitástól, de semmiképp sem nevezhető irracionálisnak.

Érteni vélem, mi motiválja Zemplén ellenvetését. Úgy látja, „itt a sikerességnek [...] két felfogása keveredik”, s igaza is van, de én máshol látom a keveredést. Szerinte egy jelenség kulturálisan sikeres, *ha sokan reprodukálják*, s mivel e reprodukálás „irracionális”, ezért „a mémek terjedésének sikeressége nem jelzője vagy mércéje racionalitásuknak”. Eszerint a sikeresség az utódok számával mérhető. Kuhnál ellenben a kognitív sikeresség (anomáliák kiküszöbölése, empirikus adekvátság stb.) *megelőzi* az utánzást: *ha* egy megoldás kognitíve sikeres, *akkor* sokan utánozzák. Ugyanez a különbség jelenik meg a szupraindividuális szelekcióról szóló bekezdésben is, de kiegészül a *kognitív sikeresség* és a *naió realizmus* közötti választóvonal elmosódásával. Itt az érv az, hogy – mivel

<sup>6</sup> Kuhn 2000. 97. – kiemelés L. J.

<sup>7</sup> Kuhn 1990b. 251.

magához a valósághoz nem férünk hozzá – adott helyen és időben nem lehet eldönteni, melyik elmélet sikeres. „A versengésben fennmaradó elmélet lesz a legsikeresebb (hiszen ezt támogatják a leginkább), de nem a végső sikeresség miatt választódik ki”, mindig csak a lokálisan leginkább sikeresnek tűnő lehet kiválasztani, s ezért „nem lehet teleologikus a szelekciós magyarázat, ha darwini akar lenni”.

A teleológia mellőzésében egyetértek, de az, hogy a tudós nem fér hozzá magához a valósághoz, nem akadályozza a kognitív hatékonyság növekedésének megállapítását. A fejlődés, mondja Kuhn, „hátról vezérelt”: mindig az elismert anomáliáktól akarnak megszabadulni, s a tudósok erre irányuló javaslatokkal állnak elő, melyek versenyéből az adott szituációban legalkalmasabb választódik ki. Ez a pragmatikus hatékonyság jól megállapítható lokálisan, s nem igényli annak ismeretét, hogy milyen „a tapasztalatokban, mérési eredményekben megjelenő túli valóság”. Mivel a sikeresség mércéje nem az *igazság*, hanem a *hatékonyság*, s ez azonnal megmutatkozik, a választás lokális, de teljesen racionális.<sup>8</sup>

A kérdés az, hogy adott történeti szituációkban leghatékonyabb elméletek ismétlődő kiválasztódásából kialakul-e a tudomány hosszú távú optimalizációja. Zemplén nemet mond erre, mivel abból, hogy a hatékonyság növekedése fogalmi diverzifikálódással jár, arra következtet, hogy a „képtelenség más tudósközösségek teljes megértésére” „szükséges a fejlődéshez”. Szerintem érdemes komolyan venni azt a különbséget, amelyet Kuhn az összemérhetetlenségből eredő *fordíthatatlanság* (nem feleltethetők meg egymásnak a különböző struktúrájú kategória-rendszerek) és a más lexikonok *megtanulásának* és *megértésének képessége* között lát.<sup>9</sup> A történész képes korábbi paradigmákban kifejezett vélekedéseket saját, a régivel összemérhetetlen lexikona segítségével újrafogalmazni.<sup>10</sup> Nem szinonim párokba állítja a két lexikon elemeit, hanem megtanulja a rendelkezésre álló empirikus adatokat a másik taxonómia segítségével rendezni és értelmezni. Ily módon kétnyelvűvé válik, s amikor a fogalmakat használja „min-

<sup>8</sup> Zemplén – mint mondja – nem látja, „milyen érvek alapján állítja Laki, hogy a \*leghatékonyabb\* eszközök választódnak ki.” Az érvek számtalan helyen szerepelnek Kuhn-nál, ezért elégséges felidézni őket: (1) „...whether or not individual practitioners are aware of it, they are trained to and rewarded for solving intricate puzzles” (Kuhn 1990b. 251.) (2) „Competition between segments of the scientific community is the only historical process that ever actually results in the rejection of one previously accepted theory or in the adoption of another” (SSR 8.) (3) A kognitív teljesítőképeségek *összehasonlítása* dönti el a versenyt (lásd például SSR 145). Ebbe az irányba kell haladnia a verifikációról folyó vitának, s ebben az értelemben értendő, hogy „Verification is like natural selection: *it picks out the most viable among the actual alternatives in a particular historical situation*” (SSR 146 – kiemelés L. J.).

<sup>9</sup> „Az összemérhetetlenség egyfajta *fordíthatatlanság* [untranslatability], ami azokra a területekre lokalizált, melyeknél két lexikai taxonómia különbözik.” E különbségek szemantikai követelmények megszegéséből erednek, de „nem akadályozzák a *közösségek közötti megértést* [intercommunity understanding].” Kuhn 1990a. 93.

<sup>10</sup> „restate past beliefs”, Kuhn 1990a. 99–100.

dig észben kell tartania, hogy épp melyik lexikon érvényes”.<sup>11</sup> Ugyanaz a hangsor az egyik lexikai keretben rendelkezik a racionális állíthatóság és igaz/hamis jelöltség tulajdonságával, vagyis teljes jogú *állítást* képez, a másikban viszont csupán fiktív, de *érthető mondat*ként szerepel, így a fogalmi diverzifikálódásból nem következik a megértés ellehetetlenülése. Egy ezzel ellentétes szemantika mégiscsak furcsa lett volna egy történéstől. Ennek szinkrón vetülete, hogy az érthetlenség lehetetlenné tenné az összehasonlítást, s akkor valóban nem lehetne látni, hogyan választódik ki a leghatékonyabb kognitív eszkörendszer.

Kuhn szerint „le kell mondani arról az elképzelésről, hogy a tudomány egységes módszer által irányított, monolit vállalkozás. Inkább úgy kell rá tekinteni, mint különféle jelenségsoportokat vizsgáló specializációk vagy speciestek sokféleségéből összeálló komplex, nem szisztematikus struktúrára.”<sup>12</sup> Ez a gondolat fejeződik ki a több írásban is visszatérő „evolúciós fa” hasonlatban: a természetfilozófia törzséről leváló diszciplínák egyre inkább elágaznak, ahogy speciális módszereket és fogalmakat fejlesztenek ki, de a fa egészére tekintve kétségtelenül egyirányú, visszafordíthatatlan fejlődést látunk. „Nagyon valószínű, hogy a lexikai diverzitásból fakadó specializáció teszi lehetővé a *tudomány egésze* [*sciences, viewed collectively*] számára, hogy a *természeti jelenségek szélesebb köréből származó rejtvényeket oldjon meg*, mint egy lexikálisan homogén tudomány.”<sup>13</sup> Ez ellentmondani látszik Zemplén értelmezésének. Az, hogy „a lokális optimalizálási folyamat [...] egyben a tudomány fragmentálódását is jelenti”, egyáltalán nincs ellentétben azzal, hogy „globális optimalizációs folyamatnak tekint[jük] a tudomány fejlődését”.

Nincs szó arról, hogy Kuhn a fejlődést célra irányuló folyamatként írja le, az optimalizáció pragmatikus értelemben veendő: vannak állandó tudományos értékek (empirikus adekvátság, rejtvényfejtő képesség, növekvő érvényességi kör stb.), s az egymást követő elméletek ezeknek egyre jobb realizálásai. Ahol a tudomány értékei meggyökereztek, „magyarázzák a természet empirikusan adekvát, konzisztens, átfogó és egyszerű leírására használatos egyre kifinomultabb – s egyre specializáltabb – eszközök folyamatos létrejöttét”.<sup>14</sup> Az evolúció mind mélységi, mind szélességi értelemben a kognitív eszkörendszer optimalizációja.

Az optimalizáció szorosan összefügg a mentális modul működésével. Zemplén ellentmondást lát abban, hogy a kuhni mentális modul „valamennyi ember számára azonos”, s ugyanakkor vannak „különböző mentális modulokat kialakító közösségek”. Úgy véli, Kuhn nem konzekvens módon használja e fogal-

<sup>11</sup> Kuhn 1990a. 100.

<sup>12</sup> Kuhn 1991. 119.

<sup>13</sup> Kuhn 1990a. 99. – kiemelés L. J.

<sup>14</sup> Kuhn 1991. 118.

mat, s ezért szükség lenne annak kritikai vizsgálatára, hogy „filogenetikailag mi humánspecifikus”, s hogyan viszonyul egymáshoz a taxonimikus és mentális modul.

Általában nem könnyű a Kuhn által használt fogalmak definiálása, de ez esetben nem látszik lehetetlennek. A vonatkozó szövegekből az az értelmezés bontakozik ki, hogy a modul két részből áll: egy neurofiziológiailag meghatározott, általános hardverből (bizonyos fizikai jellemzőkkel bíró ingerek detektálására és feldolgozására képes észlelőapparátus), valamint egy ennek konkrét működését kulturálisan beállító szoftverből. Modulról tehát kettős értelemben beszélhetünk:

- a) „Humánspecifikus” *neurális modul*. Jóllehet ugyanabban a környezetben él a denevér, a mokaszinkígyó és az ember, különbözik a számukra lehetséges tapasztalat, mivel egészen más ingerek felvételére alkalmas az észlelőapparátusuk.
- b) A közös neurális modul kulturálisan modifikált változatai a *mentális modulok*, amelyek ugyanabból az ingeranyagból eltérő észleleteket állítanak elő.<sup>15</sup>

A koronként és kultúránként különböző, hatékonyság alapján optimalizált modulok nem külön egységei a mentális architektonikának, hanem a meglévő hardver működési módjai, így a mentális modul ellentmondás veszélye nélkül tekinthető „humánspecifikusnak” és ugyanakkor közösségenként különbözőnek. A neurális modul kezelhetetlenül sok inger közvetítésére képes, s ez a túlélés szempontjából hátrányos, mivel a cselekvési alternatívák áttekintése túl sok időt vesz igénybe, illetve túl sok intellektuális kapacitást köt le. A tudományban ez a pre-paradigmatikus állapotnak felel meg: a különféle empirikus adatok, megoldandó problémák stb. egyforma jelentőségűek, ezért fókuszált kutatás helyett iskolák állandó harca zajlik. Ahogy a mentális modul szűri és jelentőségük szerint csoportosítja a beérkező ingereket, úgy a paradigma, elvégzendő kísérleteket és megfejtendő rejtvényeket jelölve ki, fókuszálttá teszi a kutatást.

A tapasztalat kognitív strukturálása fogalmi jellegű, de a használt fogalmak létezhetnek a szótlán szokások, a jártasság, a gyakorlottság szintjén, anélkül, hogy explicit neveket rendelnénk hozzájuk. Ha ez mégis megtörténik, az persze nagyban segíti a benyomások rendezését, a kialakított fajták azonosítását, de nem eredményez külön modult: a fajták elkülönítése, a hozzájuk sorolt egyedek megkülönböztetési és azonosítási képességének kialakítása, valamint az e fajtákat megnevező szavak hozzájuk kapcsolása *egyazon folyamat két oldala*. Ennek

<sup>15</sup> „A lexikai struktúra bizonyos aspektusai *biológiailag meghatározottak*, a közös filogenezis eredményei. De legalább a fejlettebb élőlények (nemcsak a nyelvvel bírók) között fontos aspektusokat határoz meg a *képzés*, a *szocializáció* [...] Biológiailag azonos lények tapasztalhatják itt-ott nagyon különbözőképpen strukturált lexikonokon keresztül a valóságot...” Kuhn 1990a. 101 – kiemelés L. J.

megfelelően, a mentális és taxonomikus modul azonos: a szelektálás és rendezés szocializációval elsajátított készsége megmaradhat nyelv előtti formában, de tesztet is ölthet egy lexikonban.<sup>16</sup>

## II. KARTEZIÁNUS SZÍNHÁZ VERSUS MEREV DESZIGNÁCIÓ

Mint az optimalizációval kapcsolatos gondolatmenetből látható, Kuhn pragmatizmusa nem annak elismerésében áll, hogy különböző problémák megoldására különböző elméletek jönnek létre, hanem abban, hogy a paradigmajelöltek közötti választásnál a kognitív hatékonyság, s nem a korrespondenciaként értelmezett igazság a mérce. Forrai Gábor ezzel kapcsolatos ellenvetése figyelmen kívül hagyja, hogy a tudomány és a valóság közötti kapcsolat *kétszintű*: a paradigmák, mint eszközrendszerek, a valóság nyomására formálódnak. Metafizikai feltevésrendszerek, módszertani standardok és konceptuális sémák között kell választani, azon az alapon, hogy melyik képes a rendelkezésre álló tapasztalatokat leginkább asszimilálni, rejtvényeket találni és megfejteni, sikeres magyarázatokat és előrejelzéseket megfogalmazni.<sup>17</sup> Ezért a „valóság” és a mentális modul által formált „világ” között van kapcsolat, mégpedig leszármazási: *a világ a valóság ismert elemeinek szelektálásával és rendezésével jön létre*. Minthogy a paradigmák az adaptálódás eszközei, a velük kapcsolatos pragmatizmus nem is lehet független az empirikus adekvátságtól, így Forrai érve célt téveszt.

Úgy látom, a realizmussal kapcsolatos ellenvetései jórészt abból fakadnak, hogy a tapasztalat önálló episztemikus szerepét teljesen megszüntető hatalmat tulajdonít az elméletnek. Bármennyire minimalista volt is azonban, Kuhn ennél realistább volt: „a Kuhn- és Laki-féle felfogásban – írja Forrai – a valóság maga *nem játszik szerepet* a referencia meghatározásában, *egyedül* mentális moduljaink és nyelvhasználati szokásaink számítanak...” (kiemelés L. J.). A mentális modul eszerint karteziánus színházat hoz létre: a paradigma fátylat képez a megismerő és a valóság között, s „a tények, melyek kijelentéseinket igazgá teszik, a mentális moduljaink által preformált világhoz tartoznak, nem [...] a független realitáshoz”. Ha így van, akkor lehetetlen a paradigmák összehasonlítása, episztemikusan értelmezhető értékkülönbségek nincsenek. Forrai meg is fogalmazza az ebből következő történeti relativizmust: „ha kialakítjuk azt a mentális modult, melynek révén newtoni erőt észlelünk, akkor ezzel immár nem észlelhetjük az arisztotelészi erőt. *Nem többet* vagy *jobban* látunk, hanem *mást* látunk” (kiemelés L. J.).

<sup>16</sup> „A következőkben [a mentális modulról] gyakran úgy beszélek, mint lexikonról, arról a modulról, melyben egy nyelvközösség tagjai a közösség fajtaneveit tárolják” – Kuhn 1990b. 229.

<sup>17</sup> Kuhn 1990b. 244.

Kuhn ezzel szemben egyirányú fejlődési folyamatnak látja a tudomány történetét, melynek egymásra következő szakaszaiban a tudósok „többet” és „jobban” látnak, mint elődeik. Galilei nemcsak „mást” látott, amikor tarthatatlannak tekintette a szub- és szupralunáris szféra kvalitatív különbségét, hanem „többet”, mint Arisztotelész. Tudniillik a Hold hegyeit és a napfoltokat, s ez az ellen a feltevés ellen szólt, hogy az égitestek más anyagból vannak, mint a földiek, ezért más mozgástörvények érvényesek rájuk. Ugyanez a helyzet a Jupiter holdjaival, amelyek szükségessé teszik az égitesteket hordozó kristályszférák valamilyen átértelmezését és így tovább. Kuhn tudománytörténeti példái azt illusztrálják, hogy a későbbi paradigmák „jobbak”, mint a korábbiak, mivel a valóság több tényét képesek tekintetbe venni. „Jóllehet sem Priestley, sem Lavoisier elmélete [...] nem volt tökéletes összhangban a rendelkezésre álló megfigyelésekkel, egy évtized múltán kevés kortárs vonakodott elismerni, hogy kettejük közül Lavoisier elmélete biztosított *jobb megfelelést*.”<sup>18</sup>

A valóság kemény feltételeket szab, melyekhez az elméleteknek igazodniuk kell. Pléh figyelmeztet nem is kerüli el, hogy a Piaget pszichológiáját jól ismerő Kuhn megtalálta annak módját, miként szabadulhatunk a felülről lefelé irányuló információfeldolgozás kizárólagosságától. A megismerés adaptáció, melynek egyik formája az inputok asszimilálása, de a másik az *akkomodáció*, a kategóriák hozzáigazítása a környezeti feltételekhez. Forrai csak az asszimiláció mozzanatát látja, s azt is meglehetősen egyoldalúan: úgy tekint az ingerekre, mint az érzékszerveinket afficiáló hatásokra, melyekből a valóságra nem is hasonlító jelenségvilágot konstruálunk. Kuhn észlelésfelfogása ellenben realiztikus: az empirikus adatok nem egyes szám első személyű, introspektíve hozzáférhető ideák, hanem *interszjektívek*. Az észlelhető tulajdonságokat mindenki képes azonosan észlelni, hiszen csakis így lehetséges az *összehasonlítás*, a *verseny* és az *osztényio általi tanítás*.

Mint a könyv záró részében igyekeztem kifejtetni, Kuhn tudományfejlődés-konceptiója kifejezetten az akkomodáció lehetőségére épül: a normál tudomány a természet rejtett tulajdonságait keresi, melyek ismételten szükségessé teszik az észlelési kategóriák, elméletek és módszerek finomítását. Azaz a valóság visszajelzésein alapuló akkomodáció nemcsak *lehetséges*, de Kuhn fontos *funkciót* tulajdonít neki: erre épül a tudomány egyirányú fejlődéséről kialakított elgondolása, az optimalizáció. Emellett, észlelőapparátusunk evolúciós fejlemény, s ha az észlelés teljesen felülről meghatározott lenne, az észleletek nem a valóság számunkra előnyös vagy veszélyes jellemzőit mutatnák, lehetetlen lenne a túlélés. A környezet „*merev korlátokat*” állít az individuális és társas élet számára, „fönmaradásunk attól függ, hogyan adaptálódunk hozzá. [...] Mi egyebet kívánhatnánk józan ésszel még egy *valóságos világtól*?”<sup>19</sup> – teszi föl a kérdést Kuhn.

<sup>18</sup> „...Lavoisier's theory provided the *better fit* of the two” (SSR 147. – kiemelés L. J.)

<sup>19</sup> Kuhn 1990a. 102. – kiemelés L. J.

A sokat emlegetett, de egyértelműen nem definiált „elméletterheltség” nem lehet azonos az empirikus-teoretikus dichotómia teljes eltörlésével.

Individuális objektumokat és azok tulajdonságait a kuhni példák szerint képesek vagyunk rámutatással azonosítani. Egy égitest, a vízimadarak csőrének formája vagy nyakának hossza ugyanúgy taxonómiától függetlenül megfigyelhető, mint a hevített fémek súlygyarapodása, láncre kötött kövek lengésideje vagy a megdörzsölt borostyán vonzó- és taszítóképesége. A tapasztalatok abban az értelemben konstruáltak, hogy szükséges a benyomások kaotikus áradásának kategoriális tagolása. A megfigyelhető tulajdonságok némelyikét „lényeginek”, másokat „akcidentálisnak” minősítünk, s az azonos lényegi tulajdonságokat mutató egyedeket egy fajtának tekintjük. E taxonómiák elmefüggőek, de nem konstruálják a *tényleges tapasztalatokat*, hanem a benyomások káoszának alkalmas szelektálásával és rendezésével megnyitják „a számunkra felfogható, lehetséges tapasztalatok végtelen körét, melyek előfordulhatnak abban az aktuális világban, melyhez hozzáférést biztosítanak”.<sup>20</sup> Azt, hogy ténylegesen milyen tapasztalatok jelentkeznek e világban, „a köznapi, és a tudományos gyakorlatot jellemző szisztematikusabb és szofisztikáltabb tapasztalásból kell megtanulni. [...] mindkettő szigorú tanár [...], s ami a nekik szentelt tiszteletteljes figyelemből származik, az a *természetre vonatkozó tudás [knowledge of nature]*”.<sup>21</sup>

Összhangban van ezzel, hogy a lehetséges paradigmajelöltek közötti választásnál alkalmazott értékek egyike az *empirikus adekvátság* (Kuhn „akkurátusságnak”<sup>22</sup> nevezi). A javasolt elmélet lehet ellentmondásban az ismert kísérleti és megfigyelési adatokkal, megfelelhet inkább vagy kevésbé nekik, vagyis ezek nem lehetnek kizárólag a mentális modul kreálmányai. Gondolhatnánk, hogy ettől még csak a Forrai-emlegette „fátyolon” jelennek meg, de Kuhn úgy tekint az efféle adatokra, mint „*magának a természetnek*” a jelzéseire, melyek képesek aláásni a korábbi elméletbe vetett bizalmat.<sup>23</sup> Márpedig, ha a helyzet az lenne, hogy a szavaink referenciájaként szolgáló entitásokat az ingereket feldolgozó mentális modul „gyártja le”, akkor nem kényszeríthetnék ki a mentális modul átalakítását.

Mint látható, Kuhn szerint a *világot* nem a semmiből vagy saját elménk pókfonalából szőjük, hanem a *valóság* rendelkezésünkre álló anyagából, s nem is szöhetjük akármilyenek. Miért különbözne ez döntően a Forrai által alternatívaként emlegetett realizmustól, amely, szavai szerint, azt jelenti, hogy a valóság „közrejátszik” a fogalmak kialakításában? Forrai kartézianus színházának fátyola elfedi az akkomodáció és az empirikus adekvátság mozzanatát, azt aényt, hogy a konceptuális sémák az *empirikus adekvátság és pragmatikus hatékonyság* minden

<sup>20</sup> Kuhn 1990b. 245. – kiemelés L. J.

<sup>21</sup> „knowledge of nature” Kuhn 1990b. 245. – kiemelés L. J.

<sup>22</sup> Kuhn 1973. 321.

<sup>23</sup> „Nature itself must first undermine professional security by making prior achievements problematic” (SSR 169. – kiemelés L. J.).



elmélethez képest külső tényei által is meghatározottak. A minimálrealizmus egyetlen fontos tekintetben tér el a Forrai által vázolttól: *nem lát elégséges okot önállóan létező fajták föltételezésére.*

A későbbi elméletek empirikusan adekvátabbak a korábbiaknál, azaz, a korábbiak rendre empirikusan tarthatatlanokká válnak, az igaznak tekintett állítások, a létezőnek vélt teoretikus entitások helyébe mások lépnek. Előbb-utóbb ez lesz a mai elméletek sorsa is,<sup>24</sup> a tudomány eddigi története nem ad elégséges alapot annak föltételezésére, hogy a valóságnak van egy metafizikai fajtastruktúrája. Amikor Putnam a konvergens realizmus föltevésével él,<sup>25</sup> dogmatikusan jár el, hisz a tudománytörténet tanúsága szerint a felgyűlő empirikus adatok rendezése újabb és újabb fajtaszerkezetek bevezetését teszi szükségessé. Kuhn szerint a metafizikai realizmus kimerül annak föltételezésében, hogy *van* egy elmefüggetlen valóság, de azt nem tudhatjuk, hogy e valóságnak van-e valamilyen szerkezete.

A görögök számos jelenséget sikeresen magyaráztak a négy elem föltételezésével, de ezek közül ma már csak a vízzel kapcsolatban vethető föl egyáltalán a kérdés, hogy természeti fajta-e, s nem tudjuk, most megtalálta-e a kémia az eszenciális tulajdonságot, s száz év múlva is érvényes lesz a „víz = H<sub>2</sub>O” azonosság. Nem tudjuk, a felgyűlő empirikus adatok értelmezésének meddig lesz alkalmas eszköze a mai atom- vagy molekulafogalom. Vannak arra utaló jelek, hogy ahogy a „víz = őselem” azonosság, úgy a „víz = H<sub>2</sub>O” is túlságosan elnagyolt definíció. A mai kémiában a protonszám azonosítja az elemeket, de vannak az adott elemre jellemző protonszámú, de eltérő neutronszámú izotópok is, melyeket nem tekintünk külön fajtának, jóllehet kémiai tulajdonságaik lehetnek meglehetősen különbözők. Az eltérő izotópokból összeálló molekulák mélyszerkezete különböző,<sup>26</sup> így az a definíció, hogy „ez és a vele azonos természetű minták” alkotják a víz természeti fajtát, finoman szólva elnagyolt. A pontosabb meghatározás legalább 18-féle természeti fajtát különböztetne meg csak a víz esetében (s akkor még nem beszéltünk az izomerekről<sup>27</sup> vagy az ionizált atomokról).<sup>28</sup>

<sup>24</sup> A Laudantól származó, s Putnam által is tárgyalt „pesszimista meta-indukció” évről van szó.

<sup>25</sup> „better descriptions of *the same entities* that the earlier theories referred to”, Putnam 1975b. 237 – kiemelés L. J.

<sup>26</sup> Az oxigénnek 3 stabil és 10 instabil izotópja van, a hidrogénnek 2 + 1. A vízmolekulák mélyszerkezete lehet H<sub>2</sub>O, D<sub>2</sub>O, T<sub>2</sub>O, HDO, HTO, DTO, + mindegyik oxigénje lehet O<sup>16</sup> vagy O<sup>17</sup> vagy O<sup>18</sup>.

<sup>27</sup> Azok a vegyületek, melyekben az egy molekulán belüli atomok sorrendje vagy térbeli elhelyezkedése különböző. Az ilyen molekulák képlete nem különbözik, de a szó szigorú értelmében vett molekulaszervezetek igen.

<sup>28</sup> Ezekben nem egyezik meg a protonok és elektronok száma, így a töltések nem egyenlítik ki egymást, s az atomnak pozitív vagy negatív töltése van. Az ionizált állapotú atomokat is tartalmazó vegyületek tulajdonsága nem pontosan azonos azokéval, melyekben nincsenek ionizált atomok.

A „természeti fajta” meglehetősen laza fogalom, mely csak akkor tesz rendet a tapasztalati sokféleségben, ha pragmatikus alapon bizonyos különbségeket figyelmen kívül hagyunk. Amennyiben viszont a  $D_2O$ -t elfogadjuk víznek, miért ne fogadhatnának el az XYZ-t is? Azért, mondhatnánk, mert az, hogy a közönséges hidrogén helyett, annak izotópja, a deutérium szerepel egy molekulában, nem eredményez esszenciális különbséget, míg az XYZ mélyszerkezet igen. De hogyan teszünk különbséget az *esszenciális és nem esszenciális tulajdonságok* között? A tudományos elméletekben játszott szerep a kritérium: a mélyszerkezeti tulajdonságok magyarázzák a felszínieket. Csakhogy épp Putnam az, aki ikerföld-példájában ezt a kapcsolatot megszünteti: a  $H_2O$  és XYZ mélyszerkezetű folyadék felszíni tulajdonságai teljesen azonosak. Ha viszont nincs *magyarázati primátusuk*, akkor a  $H_2O$  és XYZ tulajdonságok nem esszenciálisak, s ha nem azok, akkor nem is különítenek el két természeti fajtát. Összefoglalva: úgy vélem, Forrai nem jó helyen keresi a különbséget, amikor azt hangsúlyozza, hogy Putnam szerint a valóság közrehat a reprezentációknál, míg Kuhn szerint nem, hiszen minden egymástól az összemérhetetlenségig különböző taxonómiában egyaránt a realitás jelenik meg. De a realitás túl komplex, a hozzáférhető információk mennyisége pedig túl kevés ahhoz, hogy képesek legyünk egyetlen, történetileg korrigálódó taxonómiát kialakítani.

Forrai mégis látni vél egy kerülőutat: a „realizmus – állítja – [...] *abból a doktrínából táplálkozik*, melyet [...] szemantikai externalizmusnak szokás nevezni” (kiemelés L. J.). Szerintem ez a reláció épp fordított: ha találunk okot a realizmus elfogadására, akkor kialakíthatunk egy externalista szemantikát, s az eddigiekben igyekeztem megmutatni, hogy ami a természeti fajtákat illeti, nincs elégséges okunk. Forrai szerint, az externalista szemantika problémája egyáltalán nem kapcsolandó össze a kuhni minimálrealizmussal, mivel az történetileg nem a nyelvi összemérhetetlenségre való reakció volt.

Önmagában érdektelen lenne ez a filológiai kérdés, de a jelen összefüggésben sokat segít annak fölidézése, hogy a klasszikusnak számító, 1975-ös „The meaning of »meaning«” előtt Putnam több előadásban és tanulmányban<sup>29</sup> érintette már azt a kérdést, hogy miként biztosítható az elméleti keret változása ellenére a referencia változatlansága. Ezekben az írásokban egyértelműen a Kuhn és Feyerabend által fölvetett jelentésváltozás-problémával birkózott, s a megoldást 1965-től kezdve a fregei értelem és referencia elkülönítésében látta.<sup>30</sup> A referenciális diszkontinuitással szemben fogalmazott meg érveket, amelyek hol Shapere-nek a referenciájukat az elméletek változásain keresztül is megőrző

<sup>29</sup> Putnam 1965, 1973 és 1974.

<sup>30</sup> „Mindaddig, míg a »hőmérséklet« szót ugyanazon fizikai mennyiségre referálóként használjuk, akkor sem mondjuk, hogy a szó »jelentése« megváltozott, ha sokszor módosítjuk arra vonatkozó vélekedéseinket, hogy e mennyiséget pontosan milyen törvényszerűségek határozzák meg...” Putnam 1965. 128.

„transz-teoretikus” terminusaira,<sup>31</sup> hol Schefflernek arra az elgondolására hivatkoznak, amely szerint „ha valami egy dologgal kapcsolatban igaz, az független attól, hogy az objektumot hogyan deszignáljuk”.<sup>32</sup> Ahogy e szerzők Kuhnmal, ő főként Feyerabenddel vitatkozott, aki tagadta, hogy „vannak *ugyanazokra* a dolgokra vonatkozó egymást követő elméletek” s ezért nem ismerte el, hogy az olyan terminusok, mint a „hő”, „elektromosság”, „elektron” „referálhatnak különböző elméletekben ugyanarra”.<sup>33</sup> Ugyancsak vele szemben próbálta biztosítani az elméletek „kommenzurabilitását”,<sup>34</sup> így nem látszik légből kapottnak, hogy a szemantikai externalizmus kérdése tárgyalható az inkommenzurabilitással összefüggésben.

Putnam javaslata az, hogy küszöböljük ki az elméletekkel együtt radikálisan megváltozó leírásokat, s közvetlenül rendeljünk a terminusokhoz jelöletet. A közvetlen azonosítás kontextuális, mely a jelentésbe indexikus komponenst épít. A későbbi használókat kauzális-történeti láncolat köti össze a névadóval, így a név mereven deszignál. Ha például azt mondom, „Sarkozy, a jelenlegi francia király...”, akkor annak ellenére, hogy hamis leírást használok, a kifejezés azonosít egy meghatározott személyt. A példa azért tér el a hasonló russellitől, mert az individuumnév indexikus azonosító ereje mellékessé teszi a leírás hamisságát. Ezt az indexikus jelleget terjeszti ki Putnam a természeti fajták neveire.<sup>35</sup> Veszünk egy kontextuálisan azonosított vízmintát, s azt mondjuk: *ezt*, és minden vele azonos természetű folyadékot „víznek” nevezünk. Azt, hogy ténylegesen mely anyagok azonos természetűek a konkrét mintával, a megnevezés pillanatában nem kell tudni, azt a tudomány tárja föl. Ennek során ideiglenesen téves azonosításokat is elfogadnak, de ez nem érinti azt, hogy a „víz” fajtanév magára a víz természeti fajtára referál. Ahogyan a „Sarkozy” név referenciája kötött, ugyanúgy a „denevér” fajtanév is. E referenciát nem az egyes beszélők által igaznak tekintett leírások jelölik ki, hanem a tudósok.

Természetesen létezhet olyan beszélő, aki a „gólya-vagy-denevér” referenciális intencióval próbálja használni a „denevér” szót, de ez nem fog sikerülni. Putnam szerint a nyelv olyan, mint egy nagy gőzhajó, melyet egy ember nem, csak nagyobb közösség képes működtetni, ezért a nyelvközösség többi tagja szá-

<sup>31</sup> Putnam, 1973. 197. vö. Shapere 1969. A megfigyelési-elméleti dichotómia ellen Kuhn és Feyerabend által intézett támadások következményeit kivédendő, Shapere különbséget tesz az *entitások létezését kimondó* és az azokat *leíró* fogalmak között. Egy dolog, hogy bizonyos „entitások léteznek vagy nem léteznek”, illetve hogy „bizonyos problémák megoldása érdekében hogyan lehetséges vagy alkalmas kezelni őket, annak ellenére, hogy a fizikából tudjuk, valójában nem lehetnek olyanok” (157). Az „entitás” és „idealizáció” ilyen megkülönböztetése teszi lehetővé azonos entitásokról (habár ezek különböző idealizációról) szóló, versengő elméletek összehasonlítását.

<sup>32</sup> Scheffler 1967. 57–58.

<sup>33</sup> Putnam 1973. 197.

<sup>34</sup> Putnam 1974. 281.

<sup>35</sup> Putnam 1975b. 229. skk.

mára e szó továbbra is arra az állatra referál, melyet a biológusok azonosítanak. Az egyéni intenciók nem számítanak: két azonos mentális állapotban lévő és azonos szavakat használó ember referálhat különböző természeti fajtákra (víz – i-víz), illetve különböző mentális állapotban lévők azonos dolgokra (szil- és bükkfa). Mindkét eset azért lehetséges, mert a külvilág (a természet, illetve a társadalom) határozza meg a referenciát, s ez független attól, hogy valakinek a fejében téves elméletek kavarnak.

Úgy látom, a szavak jelentésének és a beszélők referenciális intenciójának elkülönülését Forrai is lehetségesnek tartja, hiszen, „az interpretáció szokásos módjának” nevezi azt az eljárást, amely egy „episztemikusan hátrányos helyzetű közösség” tagjainak szavait úgy értelmezi, hogy azok „*magukra a fátylon túli dolgokra*”, a szóban forgó beszélők számára ismeretlen „(H<sub>2</sub>O-ra)”, nem pedig a fejükben lévő elmélet által azonosított folyadékra referálnak. Ehhez még hozzáteszi, hogy saját szavaink esetében sem szabad kivételt tennünk, azok ugyanígy értelmezendők. Ha így van, akkor nehezen érthető, miért elfogadhatatlan számára az a tétel, hogy „a szavak csak [a valóság] elmeire referálhatnak, akárhol is használják [azokat] a beszélők”.

„Ugyan miért fogadná el ezt a realista?” – teszi föl a kérdést. Ez azon pontok egyike, melyeknél a „realizmus” és „szemantikai externalizmus” szinonimaként való kezelése zavart okoz. A realistának nem kell szükségképp ilyen szemantikát választania, viszont a nyelvi összemérhetetlenséggel küzdő externalista számára éppen ezt a megoldást javasolja Putnam. Ha a merev deszignáció kialakulásának imént vázolt történeti kontextusát tekintetbe vesszük, nagyon is plauzibilisnek látszik az az értelmezés, amelyet Forrai elvet: Putnam célja valóban „annak megakadályozása, hogy a tudomány forradalmi változásai során megváltozhasson a terminusok referenciája”. Annyit pontosítanak csupán e tételen, hogy ez nem a *realista*, hanem a *szemantikai externalista* célja, bár utóbbi elképzelése összefügg a realizmussal. Lássuk, hogyan!

A Forrai által kifogásolt másik tézisem az volt, hogy a „valóságnak van egy adott, igazi metafizikai szerkezete”. Úgy látom, egy mérsékelt realizmus fenntartható, ha lemondunk is e föltevésről, de az externalizmus csak akkor, ha szemantikánkat nem akarjuk a fajtanevekre kiterjeszteni. Ezek esetében az externalizmus azt jelenti, hogy vannak tőlünk független természeti fajták, melyekhez a neveket kapcsoljuk. Ennek hiányában a fajtanevek referenciáját semmi sem stabilizálná, különböző korok, sőt egyes emberek különféle jelentésekkel használnák. Putnam szerint az esszencia meghatározza, „mit jelent *víznek lenni bármely lehetséges világban*”.<sup>36</sup> Ha nem feltételezzük, hogy a fajták metafizikailag kötöttek, milyen alapon állítjuk, hogy a földi és ikerföldi emberek más-más természeti fajtáról beszélnek, anélkül, hogy bármiféle különbségről tudnának?

<sup>36</sup> Putnam 1978. 73–74.

Nekem úgy tűnik föl, ilyesmiről beszél Forrai is, mikor azt mondja, „a realizmust [...] az különbözteti meg a minimálrealizmustól, hogy úgy tartja: amennyiben sikerül referálnunk, akkor *valóságos dolgokra* referálunk” (kiemelés L. J.). Nem azt kell ezen érteni, hogy a „referál” ún. „siker-ige”, amelyet akkor használunk jogosan, ha nemcsak úgy tűnik, hogy referáltunk valamire, hanem valóban ez történt? Ha nincs mire referálni, akkor nem történik referálás, csak annak látszata, vagyis ha nem léteznek a természeti fajták, akkor a fajtanevek egyáltalán nem referálnak. Putnam referenciaelmélete nyilván nem az, hogy a beszélők „úgy vélik, hogy referálnak”, hanem az, hogy attól függetlenül, hogy mit vélnék, ténylegesen referálnak, így a putnami externalizmus implikálja a fajták elmefüggetlen létezését. Ha nem így van, akkor nincs értelme azt mondani, hogy a „víz” hangsor valami „a fátylon túli dologra”, azaz egy természeti fajtára referál, függetlenül attól, hogy a beszélőnek van-e fogalma a kémiai atomelméletről.

Az externalista jelentéselmélet fajtanevekre való kiterjesztésével szemben három ellenvetést hoztam föl:

1. Nincs elégséges alapunk föltételezni, hogy léteznek természeti fajták, azaz, hogy a fajtanevek referenciája externálisan meghatározott.
2. Maga a fajta mint absztrakt entitás nem azonosítható közvetlenül, elkerülhetetlen egy elmélet bevezetése.
3. Az osztenzíve kijelölt minták nem szükségképpen tartoznak egyazon fajtához.

Az első pont, mint láttuk, a pesszimista meta-indukció következménye, a második abból következik, hogy fajtákkal mint absztrakt entításokkal nem kerülhetünk kauzális kapcsolatba. Azt, hogy egy konkrét dolog valamely fajta eleme-e, a tudomány dönti el, azaz a fajtanevek referenciájának kijelöléséhez be kell vezetni egy az indexikus *nevet* („ez”) kiegészítő *leírást* („ugyanilyen természetű”), s ez teoretikus elemet tesz a jelentéshez. Ami a harmadikat illeti, egy-egy vízmolekula, mint láttuk, attól függően, hogy a H és O-izotópok milyen kombinációjából áll, meglehetősen sokféle lehet. De ennél még sokkal nagyobb variációs lehetőségeket engednek meg a Putnam által a természeti fajtanev referenciájának rögzítésére használt „vízminták”, amelyek a sokféle izotópból összeálló, még többféle molekulából állnak. Azaz, két beszélő nemcsak akkor mutat különböző mikroszerkezetű anyagra, ha egyikük a Földön, másikuk az Ikerföldön van, de jó eséllyel akkor is, ha mindketten a Földön vannak.

Természetesen Forrainak igaza van abban, hogy Putnam elhatárolta magát mind a metafizikai realizmustól, mind a mágikus referenciaelmélettől, de ez a 80-as évek elejére bekövetkezett „belső realista” fordulattal történt meg. Ekkor már úgy fogalmazott, hogy „nem léteznek konceptuális sémáktól független tárgyak. *Mi* szabjuk föl a világot dolgokra, amikor ilyen vagy olyan leíró sémát veze-

tünk be.”<sup>37</sup> A belső realizmus nem látszik különbözni a minimálrealizmustól, így Kuhn nem minden alap nélkül mondhatta e fordulat láttán Putnamról: „most végre az én nyelvemet beszéli”.<sup>38</sup> A valóság Kuhn szerint is „közreműködik” szavaink referenciájának kialakításában, s az így szerzett adatokat a paradigmák fajtákba rendezik, de ez nyilván nem elégséges a fajtanevek referenciájának me-revségéhez. Nem világos, hogyan tarthatja fenn Putnam a szemantikai externalizmust az elme- és elméletfüggetlen fajták létezésének metafizikai fölteve-se nélkül. Az, hogy a „realizmus” és „szemantikai externalizmus” nem szinonim, itt fontossá válik: a paradigma megváltozásával a fajtaszerkezet, azaz a fajtane-vek referenciája is megváltozik, de ez nem jelenti a realizmus teljes föladását.

A fajtanevek esetében kétségesnek látszik az externalizmus *társadalmi esz-közökkel* való biztosíthatósága is. Az individuumnevek referenciája azért merev, mert a nyelvközösség a keresztelési aktushoz visszavezető kauzális-történeti láncolatra hagyatkozik. A láncolat végpontján *kitüntetett episztemikus pozícióban lévő* és megfelelő *társadalmi fölhatalmazással rendelkező* személy áll. Episztemi-kus kitüntetettsége abból adódik, hogy számára a kontextus lehetővé teszi a megnevezett dolog kontextuális azonosítását, társadalmi fölhatalmazottsága pe-dig abban áll, hogy jogosult a névadás beszédaktusára. Ennek analógiájára van elgondolva a fajtanevek szemantikája: a nyelvhasználók túlnyomó része nem képes azonosítani az arany fajtát, de a nyelvi munkamegosztásnak köszönhe-tően, ráhagyatkozhat egy az ékszerészeken, tanárokon, ismeretterjesztőkön ke-resztül a kémikusokig/fizikusokig vezető láncolatra, melynek a végpontján állók ugyanolyan biztonsággal azonosítják a fajtanev által deszignált fajtát, mint a ke-resztelési aktus résztvevője a megnevezett individuumot.

Szerintem ez az analógia hibás. A keresztelésnél jelen lévők felhatalmazott-sága csak a *névnek a megnevezendőhöz való hozzákapcsolásra* vonatkozik, a tudó-sok esetében viszont a *megnevezendő entitás (a fajta) azonosítására* is. Ez jelentős különbség, mivel a megnevezés érvényessége társadalmi kérdés, az azonosítás azonban nem. Az individuumot elkeresztelő személy támaszkodhat helyzetéből adódó episztemikus kitüntettségére, a fajtát azonosító tudósok azonban *nin-csenek ilyen kitüntetett pozícióban*. Legfőljebb ideiglenes konszenzust alakítanak ki, s még abban sem lehetnek biztosak, hogy van valami, amit megneveznek, a valóság nem kontinuum spektrum. A társadalmi externalizmus tekintetében a tudósok a nyelvközösség páriái: ők ugyanis nem hagyatkozhatnak kitüntetett episztemikus pozícióban lévő keresztelőkre. A fajtanevek referenciáját sem a valóság, sem a társadalom nem fixálja, az csupán a tudósok elméjébe horgony-zott, velük vagy a konszenzusukkal együtt elmúlik, semmi nem biztosítja a le-hetséges világokon keresztüli azonosíthatóságot.

<sup>37</sup> Putnam 1981. 52.

<sup>38</sup> Kuhn 1995. 312.

Mivel a fajtanevek externális meghatározása tekintetében mind a fizikai, mind a társadalmi környezet problematikus, Putnam eredeti, a referenciális stabilitás lehetőségét kereső programja itt sikertelen. Kuhn-interpretációm azt igyekszik megmutatni, hogy ha a természeti fajtákra nem terjesztjük ki a merev deszignáció koncepcióját, az nem feltétlenül jár katasztrofális következményekkel. Az erős relativizmus elkerülésében fontos szerepet játszik az összemérhetetlen paradigmák közötti megértés lehetősége, s ez kétségkívül megnyitja a kuhni tudományfilozófiát a hermeneutika felé.

### III. TUDOMÁNYFILOZÓFIA ÉS HERMENEUTIKA

Így hát ellenérzések nélkül olvastam Schwendtner Tibor javaslatát, amely szerint Kuhn tudományfelfogását célravezetőbb lenne hermeneutikaiként értelmezni. A tudományfilozófia és hermeneutika közötti kapcsolat legalább Dilthey óta föl-fölmerülő kérdés, melynek az utóbbi két-három évtizedben komoly irodalma keletkezett. Úgy látom, érdemes megkülönböztetni általában a tudományfilozófia és a hermeneutika, valamint konkrétan, Kuhn tudományfilozófiai nézetei és a hermeneutika közötti kapcsolatot. Mindkét esetben föltehető az a kérdés, hogy *lehetséges-e* hermeneutikai-fenomenológiai keretben tárgyalni a tudományfilozófiai problémákat, s a válasz meglehetősen egyértelmű: igen. Ezzel nem is kívánok részletesen foglalkozni. Nehezebb azonban a válasz arra a kérdésre, hogy *érdemes-e* ezt a hagyományt mozgósítani az értelmezéshez.

Ami általában a tudományfilozófiát illeti, úgy érzekelem, hogy bár erre hermeneutikai oldalról komoly kísérlet történt, a két hagyomány között nem jött létre kölcsönösen megtermékenyítő együttműködés. Ellenségességről már nem beszélhetünk, inkább talán ignorálásról, melynek bizonyos okai jól láthatók.

1. A legfontosabb az, hogy a hermeneutika túlságosan gyakran választ kritikájának célpontjául egy *konstruált*, őspozitivistá tudományfelfogást. Hadd példázzam ezt egy P. Heelan-tól vett idézettel: „...a mainstream tudományfilozófiában a valóság nem kontextuális, hanem nominalisztikus, nem értékekkel átitatott, hanem értékmentes, nem történeti, hanem abszolút, nem többértelmű, hanem világos, ha észleljük, akkor csak passzívan (nem kreatívan) befogadjuk, mivel az észlelés nem aktív, történeti átalakítás, hanem a valóság tükrözése...”<sup>39</sup>

Aki ismeri a tudományfilozófia 1983-as állapotait, könnyen gondolhat a megfelelő informáltság hiányára, rosszindulattal esetleg arra, hogy itt döglött oroszlán rugdosása folyik. Mint könyvemben is igyekeztem megmutatni, a társadalmi és történeti kontextusából kiemelt, „világtalanított” megismerő szubjektumról és annak a logikával azonosított racionalitásáról

<sup>39</sup> Heelan 1983. 214.

és nyers empirizmusáról kialakított naiv képet (legkésőbb a 30-as évektől kezdődően) maga a tudományfilozófia bontotta le. Amit Heelan 1983-ban „mainstreamnek” tekint, az sokkal inkább a felvilágosodás propagandaképe, melytől már a 19. század bizonyos tudományfilozófusai (Whewell, Duhem, Mach) is távol álltak. A mai poszt-pozitivist tudományfilozófia pedig inkább úgy tekint a tudományra, mint kultúrára és emberi gyakorlatra, melynek van társadalmi közege, kulturális kontextusa és történeti dimenziója, és semmiképp sem merül ki időtlen logikai struktúrák feltárásában vagy értékmentes objektivitásában. Ha a tudományfilozófia gyakorlata és a hermeneutika víziója között ekkora a távolság, a kritika ignorálása magától értetődő: a tudományfilozófiának egyszerűen nincs dolga vele, mivel egyetértenek.

2. Az egyetértést a hermenuták is látják, de ez gyakran eredményez unalomba fúló elemzéseket, amelyek jórészt annak megmutatásában merülnek ki, hogy például Heidegger nagyon hasonló állításokat tett a tudományról, mint a poszt-pozitivist tudományfilozófusok. Az ilyen tanulmányok jelentős részben Heidegger szótárának ismertetéséből állnak, kiegészítve néhány párhuzammal: lám, már a 20-as években is olyasmiről beszélt, mint később a tudományfilozófusok. Egy ilyen könyvet *egyszer* lehet megírni, de a történeti elsőbbség kimutatása után csak annak van értelme, ha a hermeneutikai megközelítés alkalmazása előreviszi a tudományfilozófiát.
3. Ez utóbbi gyakran hiányzik a tudományfilozófiai kérdéseket tárgyul választó hermeneutikai írásokból. A gondolkodás „előzetesség-struktúrája”, a tapasztalat nyelvisége és körszerűsége egyfelől, másfelől a „paradigma”, az „elmélettel való átitatottság” vagy a „konceptuális sémák” közötti strukturális megfelelés fennáll, de mit nyerünk a hermeneutikai fogalmakra való lefordítással? Legalábbis kérdéses, mennyiben revelatív 1997-ben az a megállapítás, hogy „a tudományos leírás, magyarázat és megértés minden formája kifinomult interpretáció. A tudós nem azt állapítja meg, hogy egy dolog micsoda, hanem csupán azt, hogyan mutatkozik meg adott föltevések mellett. Így a tudós mindig sajátos fore-having, fore-sight és fore-conception-nal bíró hermeneutikai szituációban lát munkához.”<sup>40</sup>
4. Nem állítom, hogy a tudományhermeneutika egyáltalán nem hozott új gondolatokat, de érteni vélem, hogy Heelan, Kockelmans vagy Kisiel neve miért nem szerepel a tudományfilozófiai hivatkozásokban: a tudománnyal kapcsolatos igazán *új gondolatok* nem ebből az irányból jönnek, a tudománytörténeti és szociológiai vizsgálatok (STS, SSK), a tudomány antropológiai megközelítései, a kognitív pszichológia és diskurzus-analízis inkább szolgál új tényekkel és felismerésekkel. Emellett, ezek az empirikus kutatások képesek dialógust kezdeni a tudományfilozófiával, sőt a tudósokkal is, míg

<sup>40</sup> Kockelmans 1997. 312.



a teljesen más fogalmiságot használó hermeneutika nemigen. A *történeti elsőbbség* elismerése után fölmerül az *indokoltság* kérdése: érdemes-e a kissé nehézkes, „filozófiai” megközelítést szorgalmazni, ha nem vezet többre, mint olyan belátásokra, melyeket empirikus vizsgálatok sokkal konkrétabban mutatnak?

Ami a kuhni elméletet illeti, az alapkérdésben nincs vita Schwendtner és köztem: ahogy a könyvben is jeleztem, Kuhn gondolatainak van hermeneutikai dimenziója. Fontosak azonban a hangsúlyok. Schwendtner egy lábjegyzetben idézi Kuhnnek azt a megjegyzését, hogy „A hermeneutika legközvetlenebb és legdöntőbb hatása [...] a tudományról kialakított képemet érintette”, s ezzel alátámasztani véli, hogy *Kuhn egész tudományértelmezése hermeneutikai jellegű volt*. Az idézett mondat valóban ott van a szövegben, de Schwendtner saját jelentésváltoztatásával felé mozdítja el értelmét. A megjegyzés olyan összefüggésben hangzik el, amelyben Kuhn arról felismeréséről ír, hogy a távoli történeti korok tudósai más fogalmakkal írták le és értelmezték a természetet, ezért írásaik nem interpretálhatók mai fogalmakkal. Nem több ez, mint a prezentista történelemszemlélet tarthatatlanságának kimondása, ami már a TFSZ-ben is megjelent. 1962-ben még szó sincs Kuhnna hermeneutikáról, sokkal inkább az anti-whig történeti szemléletről, mellyel Koyré és Butterfield írásaiban találkozott.

Schwendtner feszültséget érez az alkalmazás elvével kapcsolatban „Kuhn tényleges gyakorlata” és az alkalmazás hermeneutikai problémájára vonatkozó interpretációm között. Hermeneutikailag naivnak minősítettem Kuhn történeti attitűdjét, de egyáltalán nem hiszem, hogy a megoldás annak bemutatása lenne, „mit mond ma [...] számunkra a szöveg” (kiemelés L. J.). Ezzel kapcsolatban Schwendtner Gadamert idézi: „a megértés [...] egyben mindig alkalmazás”,<sup>41</sup> de figyelmen kívül hagyja, hogy az adott helyen arról van szó, hogy a törvényt és a vallási szöveget *nem történetileg* megérteni, hanem *jogi érvényében konkretizálni* kell. A törvény vagy az ige önmagábanvaló értelme nem érdekes, az ítéletben és a prédikációban a ma érvényesnek tekintett értelemnek kell megjelennie. Általánosan: „a szöveget [...] minden konkrét szituációban újból és *másképp* kell érteni.”<sup>42</sup> Kuhn nem kezd dialógust a múlttal, nem azt akarja megállapítani, hogy mit mond a mai értelmezési horizontba illesztve a régi szöveg. Történeti dokumentumként, a maga másságában akarja azt rekonstruálni, s nem azt keresi, hogy milyen érvénye van számunkra. A „másképpértés” nem lehet a „tényleges gyakorlatát” irányító hermeneutikai elv, mivel Kuhnna az volt az öndefiníciója, hogy – szemben a kumulatív és kontinuos képet rajzoló hibás tudománytörténettel – ő nem konstruált történetet ír.

<sup>41</sup> Gadamer 1984. 219.

<sup>42</sup> Uo. – kiemelés L. J.

Ez indokolja, hogy szerintem Kuhn hermeneutikája nem azért tekinthető „pozitivistikusnak”, mert „a tudománytörténeti tényeknek [...] túlzó fontosságot tulajdonít”, hanem azért, mert a történeti tényeket az értelmező horizontján kívül eső, objektív adottságokként kezeli. Ebből következően, nem is lehet az a véleményem, hogy e hermeneutika „a lehetséges interpretációk sokféleségét és bizonytalanságát vitte be a tudománytörténeti értelmezésbe”. Épp ellenkezőleg: azzal, hogy a tudománytörténeti tényeket objektívnek tekinti, mellőzi a történész jelentéskonstituáló szerepét és történeti szituáltóságát, éppenséggel *egyetlen lehetséges interpretációt föltételez*. Ez némiképp váratlan, hiszen saját elméletéből az következne, hogy a történészek értelmezési horizontját ugyanúgy a mindenkori paradigma határozza meg, mint bármely tudósét. Ennek elismerése azonban relativizálná a tudománytörténetet, s megszüntetné annak a tudományképünk átalakításához szükséges normativitását. Ezért Kuhn még azon az áron is ragaszkodik az immanens történetfelfogáshoz, hogy az a paradox helyzet áll elő, hogy a tudománytörténetet úgy véli rekonstruálni „wie es eigentlich gewesen”. Ez kifejezetten szemben áll Schwendtner azon tézisével, miszerint Kuhn „alkalmazott tudományfilozófiát” művel, amelyben a történeti esettanulmányok csupán egy a priori elmélet téziseit szemléltetik.

E gondolat másik vetülete a „példázat” szerepének erős szűkítése: mi más lenne a példázat, mint konkrét probléma megoldására alkalmazott általános elmélet – kérdi Schwendtner. Azonban, ahogy a tudománytörténet tényei, ugyanúgy az exemplárok is megelőzik az elméleteket.<sup>43</sup> Emellett Kuhn példázatnak nevezi Arisztotelész *Fizikáját* vagy Newton *Principiáját* is,<sup>44</sup> noha ezekre aligha mondhatnánk, hogy valamely elmélet konkrét, egy problémára való alkalmazásai. Inkább a maguk egészében szolgálnak általános mintaként arra, hogyan kell problémákat azonosítani és megoldani. A másik végletet reprezentálja az a példázat, amellyel fizikai tanulmányai során találkozott: Newton első mozgástörvényének bevezetése az üveglapon csúszó jégtömbbel. Még szó sincs általános elmületről, egy felépítendő elmélet axiómájának intuitív megalapozása történik.<sup>45</sup>

Általánosságban azt mondhatjuk, hogy a példázatok egy gondolkodás és észlelésmódot, problémák azonosítását és bevált megoldástípusait szemléltetik és gyakoroltatják be. Szemléletmódot, kompetenciát alakítanak ki, nélkülözhetővé teszik az általános módszertani szabályokat, s velük a konkrét helyzetekre való alkalmazás szabályaival járó végtelen regresszust. És még egy alapvető, minden elméletet megelőző funkciót ellátnak: mint a hattyú – kacsá – liba fajták elkülönítésének és megnevezésének megtanulását leíró kuhni történet mutatja,

<sup>43</sup> Vö. SSR V. fejezet: A paradigmák elsőbbsége.

<sup>44</sup> SSR 10.

<sup>45</sup> Ez az exemplár a természet évezredes látásmódjának átalakítását célozza: a mozgás állapotot, melynek nem fönntartása, hanem megváltoztatása igényel erőt.

konkrét példák teszik lehetővé az empirikus benyomások özönének rendezését és a szavak valóságba horgonyzását. Ugyanígy, a külvilág és a nyelv közötti szemantikai kapcsolat megteremtése történik tudományos kontextusban is, amikor a diákok példák segítségével elsajátítják az „erő”–„súly”–„tömeg” egymást kölcsönösen definiáló fogalmait Newton III. törvényének és a Hook-törvénynek a felhasználásával.<sup>46</sup> Itt sincs szó kész elmélet alkalmazásáról, csak a newtoni fizika alapfogalmainak demonstráció segítségével végzett kialakításáról. A diákok definíciók és elméletek ismerete nélkül, közvetlenül kapcsolják<sup>47</sup> e szavakat a referenciáikhoz, egyúttal megtanulják felismerni ezeket a fizikai entitásokat a környezetükben.

Bár eszemben sincs tagadni, hogy Kuhn számos fontos gondolata hermeneutikai, nem érzem termékenynek, ha ettől az elfogadható kiinduló-állítástól elfogadhatatlan általánosításhoz lépünk tovább. Egy dolog ugyanis, hogy Kuhn-nak vannak hermeneutikai belátásai, s egészen más azt mondani, hogy „egész tudományképe legjobban e háttér előtt értelmezhető”. Nem látom tarthatónak azt az interpretációt, hogy bár a hermeneutikai hagyományt nem ismerte, „a kuhni gondolkodásmód egészét hermeneutikai jellegűnek” kellene tekinteni. Lehet, hogy „Kuhn-nak nincs olyan jelentős gondolata, melynek ne lenne a hermeneutika által megfogalmazható verziója”, maga Kuhn azonban a hermeneutikai hagyományt homályosnak érezte.<sup>48</sup> Nem tagadta fontosságát, de ténylegesen csak annyi történt, amennyit állított: a hetvenes évek elején megismerkedett a „hermeneutika” *terminussal*. Semmi okunk föltételezni, hogy ez egyúttal a *hermeneutikai hagyomány* megismerését és elfogadását jelentette, s egy későbbi előadásban azt is leszögezte, hogy „a természettudomány [...] nem hermeneutikai vállalkozás”.<sup>49</sup>

Természetesen lehet a szerzőt explicit intenciói ellenében is értelmezni, de csak akkor, ha ezzel megvilágítjuk szövege egyébként homályban maradó jelentésrétegeit. A hermeneutikai fogalmakban való újraírás nyomán azonban inkább azzal a veszéllyel szembesülünk, hogy az amúgy sem mindig egyértelmű kuhni fogalomkészletet esetleg még homályosabbra cseréljük. Csökken-e a paradigma 21 jelentéséből származó zavar, ha előzetesség-struktúráról kezdünk beszélni? Ezt a kérdést érdemes fölvetni, de a válasz számomra nem látszik biztatónak.

<sup>46</sup> Kuhn 1990c. 302–3.

<sup>47</sup> Vö. „Rather than being defined, these terms are introduced by exposure to examples of their use.” Kuhn 1990c. 302.

<sup>48</sup> „continue to find opaque” Kuhn 1977. xv. Szögezzük le: a „hermeneutika” szó 1972 után sem tartozott Kuhn szókincséhez, s ez nem arra utal, hogy a találkozást revelációként élte volna meg. Élete végén adott nagy interjújában sem tagadta, hogy bizonyos értelemben újra felfedezte a hermeneutikai szemléletmódot, de sietett hozzátenni: „...amikor azt kérdezik, nem ezt – vagy valami ilyesmit – mondott-e Heidegger is?», azt felelem, igen, meg lehet, de én nem olvastam...” Kuhn 1995. 321.

<sup>49</sup> Kuhn 1989. 222.

Mindettől persze még lehetne Kuhn hermeneuta, de nemcsak *faktuálisan* nem az, hanem több fontos vonatkozásban *koncepcionálisan* sem. Schwendtner fontos tézise, hogy Kuhn értelmezése érdekében érdemes lenne a gadamerinél „tudományfilozófiai szempontból relevánsabbnak tűnő heideggeri hermeneutikához fordulnunk”. Heidegger szoros kapcsolatot lát a teoretikus beállítódás és az ittlét világban-való-léte, a tudomány és a faktikus élet között. A tudomány a nyelviileg konstituált, morális, társadalmi, politikai, vallási értékek és intenciók által áthatott, előzetes világértésből nő ki. A köznapi világban-való-lét előzetes világértelmezést és fogalmiságot határoz meg a kutató számára, s a tudományos megismerés ennek megfelelően nem különül el élesen az értelmező átlagos köznapi létértelmezésétől.

Ezzel szemben Kuhn kifejezetten azt tekinti az érettség jelének, ha egy tudomány leválik az életvilágról. Az érett tudomány világa nem az a világ, melyben az átlagember otthonosan érzi, kiismeri magát, hanem attól izolált szféra, a laikusok számára teljesen érthetetlen nyelvvel és gyakorlattal. Ez részben a specializációból fakadó ezoterizálódás eredménye, részben azonban szándékos izoláció: a tudomány normák és értékek által el van szigetelve a környezetből származó megrendelésektől, elvárásoktól, érdekektől.<sup>50</sup> A teoretikus beállítódás nem a világban-lét átlagos mindennapiságából nő ki, hanem egy elszigetelt szubkultúra speciális diskurzusa és gyakorlata által kialakított mesterséges hagyomány. Az ebben az elszigetelt szférában zajló interakciók és értelmezések tanulmányozása inkább *empirikus*, semmint *filozófiai* vizsgálatot kíván.

Fölmerül tehát a kérdés; valóban érdemes lenne-e „egységes filozófiai keretben értelmezni Kuhn-t”, ahelyett hogy „szétcincálnánk szociológiára, pszichológiára” stb. Azt hiszem, itt alapvető felfogásbeli különbség van köztünk, hiszen e javaslat épp azt az újszerűséget szüntetné meg Kuhnban, mely a maga idejében forradalmi volt. A kuhni tudományfilozófia nem mellékesen naturalizáló, hanem épp ebből fakadnak új belátásai. Nemcsak a hermeneutikai, de semmiféle a priori filozófiai keret nem lenne megfelelő a rekonstruáláshoz, hiszen Kuhn arra törekedett, hogy a tudományt „a tényleges kutatási tevékenységből”<sup>51</sup> értelmezze. Emellett, az episztemológia naturalizálásának gondolata egyáltalán nem állt távol a kortárs kognitív forradalom bizonyos eredményeit felhasználó Kuhntól, aki többször hivatkozott arra is, hogy a pszichológiai, nyelvészeti, tudomány-, sőt művészettörténeti kutatások egyre világosabbá teszik, hogy a megismerő szubjektumról, az észlelésről, a nyelvről kialakított hagyományos filozófiai elképzelés tarthatatlan. Ebben a kérdésben teljesen igaza van Pléh Csabának, aki a neurobiológia kísérleti eredményeinek a tárgyalásba való bevonása mellett érvelve éppen azt javasolja, hogy a szaktudomány azóta föltárt tényeinek

<sup>50</sup> Ezért nem alkalmazható rá Kuhn szerint a Zemplén által hiányolt Bloor- és Barnes-féle externalista szociológia sem.

<sup>51</sup> SSR 1.

fényében értelmezzük az elmélettel való átitatottság jelenségét. Nem értek tehát egyet Schwendtner azon megállapításával, hogy „elmulasztottam” volna a kuhni gondolatok egységes filozófiai értelmezésének lehetőségét. Jobb leírása a helyzetnek, hogy szándékosan kitértem előle, mert bár *lehetségesnek* láttam, de *szükségesnek* vagy az *értelmezést segítőnek* nem.

## IRODALOM

- Gadamer, Hans-Georg 1984. *Igazság és módszer. Egy filozófiai hermeneutika vázlatja*. Ford. Bonyhai Gábor. Budapest, Gondolat.
- Heelan, P. 1983. Natural Science and Being-in-the-World. *Man and World*. 16. 207–219.
- Kockelmans, J. J. 1997. On the hermeneutical nature of modern natural science. *Man and World*. 30. 299–313.
- Kuhn, Thomas S. 1973. Objectivity, Value Judgement, and Theory Choice. In Kuhn 1977. 320–339.
- Kuhn, Thomas S. *The Essential Tension*. Chicago–London, The University of Chicago Press. 1977.
- Kuhn, Thomas S. 1989. The Natural and the Human Sciences. In Kuhn 2000. 216–223.
- Kuhn, Thomas S. 1990a. *The Road since Structure*. In Kuhn 2000. 90–104.
- Kuhn, Thomas S. 1990b. Afterwords. In Kuhn 2000. 224–252.
- Kuhn, Thomas S. 1990c. Dubbing and Redubbing: The Vulnerability of Rigid Designation. In C. W. Savage (szerk.) *Scientific Theories*. Minneapolis, University of Minnesota Press. 298–318.
- Kuhn, Thomas S. 1991. The Trouble with the Historical Philosophy of Science. In Kuhn 2000. 105–120.
- Kuhn, Thomas S. 1995. (A. Baltas – K. Gavroglu – V. Kindi) A Discussion with Thomas S. Kuhn. In Kuhn 2000. 255–323.
- Kuhn, Thomas S. 2000. *The Road since 'Structure'. Philosophical Essays, 1970–1993, with an Autobiographical Interview*. J. Conant – J. Haugeland (szerk.) Chicago – London, The University of Chicago Press.
- Putnam, Hilary 1965. How not to talk about meaning. In Putnam 1975a. 117–131.
- Putnam, Hilary 1973. Explanation and reference. In Putnam 1975a. 196–214.
- Putnam, Hilary 1974. Language and reality. In Putnam 1975a. 272–290.
- Putnam, Hilary 1975a. *Mind, Language and Reality. Philosophical Papers*. vol. II. Cambridge, Cambridge University Press.
- Putnam, Hilary 1975b. The meaning of 'meaning'. In Putnam 1975a. 215–272.
- Putnam, Hilary 1978. Reference and Truth. In Putnam 1983. 69–86.
- Putnam, Hilary 1981. Two philosophical perspectives. In Putnam 1983. 49–75.
- Putnam, Hilary 1983. *Realism and Reason. Philosophical Papers*. vol. III. Cambridge, Cambridge University Press.
- Scheffler, I. *Science and Subjectivity*. Indianapolis–New York–Kansas City, The Bobbs-Merrill Co. 1967. 57–58.
- Shapere, D. 1969. Notes toward a post-positivistic interpretation of science. In P. Achinstein – S. Barker (szerk.) *The Legacy of Logical Positivism*. Baltimore, The Johns Hopkins Press. 115–162.



MÁRTON MIKLÓS

## Philosophia perennis

Ujvári Márta, *Metafizikai dilemmák. Szubsztanciák, trópusok, változás.*  
Budapest, L'Harmattan Kiadó, 2009. 215 oldal

Micsodák valójában a körülöttünk lévő dolgok, mi az igazi természetük? Ez volt az a kérdés, amellyel a nyugati filozófia tudásunk szerint útjának indult több mint 2500 évvel ezelőtt, és ez az a kérdés, amellyel Ujvári Márta legújabb, 2009-ben megjelent könyve is foglalkozik. Itt tehát valóban örök – de legalábbis hosszú életű – filozófiai kérdésről van szó, ám a szerzőnek egyáltalában nem vette el a kedvét az a tény, hogy a preszókratikusoktól a skolasztikáig, Leibniztől Russellig már rengeteg filozófus megkísérelt választ adni a kérdésre. Ujvári könyve ugyanis nem tankönyv, nem is történeti áttekintés vagy tanulmánygyűjtemény, hanem a szerző saját álláspontját kifejtő és védelmező monográfia. Mint ilyen, egyedülálló a hazai filozófiai életben – hirtelenjében nem is tudnék hasonló magyar nyelvű vállalkozást említeni az utóbbi évekből –, és ez, úgy gondolom, önmagában értékessé teszi a művet.

Nos, Ujvári szerint az úgynevezett kontinuáló dolgok – avagy hogyan gyakran nevezi: a szubsztanciák – valójában trópusok nyalábjai. A trópusok maguk pedig tulajdonságparkulárék. Vagyis, ahogy a szerző egy helyen megfogalmazza: „a trópuselmélet kemény magja” az az állítás, hogy „a konkrét parkulárékról állítható kvalitatív vonások maguk is parkulárék” (77). Az általa képviselt elmélet lényege tehát, hogy „a konkrét szubsztanciák in-

kább tulajdonságparkuláréknak, semmint közösen birtokolható és birtokolt univerzáléknak legyenek a nyalábjai” (77). Maguk a hagyományos értelemben vett tulajdonságok pedig, amelyek egyaránt jellemezhetnek több parkulárét is, a trópusok hasonlóságosztályaival lesznek megadhatók, vagyis egy tulajdonság nem más, mint a *pontosan hasonló* trópusok halmaza (109).

Egy ilyen koncepciónak két fontos kérdésre kell választ adnia: (1) Mik is azok a trópusok?, (2) Milyen előnyei vannak a trópusnyaláb-elméletnek a rivális elméletekkel szemben? Ujvári a III. fejezetben tárgyalja a trópusok individuációjának kérdését. Itt elutasítja egyrészt a trópusoknak a bennük megjelenített tulajdonságok révén történő azonosítását – hiszen így voltaképpen univerzálék példányait kapnánk valódi parkulárék helyett –, másrészt azt a nézetet, hogy a trópusok primitív, további struktúra nélküli parkulárék lennének. Ez utóbbi nézet azért implauzibilis a szerző szerint, mivel lehetővé teszi, hogy egy időben, egy helyen két numerikusan különböző, ámde kvalitatíve megkülönböztethetetlen trópus is előforduljon. Ez nemcsak hogy kontraintuitív elgondolás, de nem szolgálja a trópusnyaláb-elmélet érdekeit sem, hiszen mi értelme volna mondjuk egy almának két pirosságot mint alkotórészt tulajdonítunk?

Ujvári e koncepciókkal szemben a trópusok azonosítását a hordozó szubsztanciára történő hivatkozással tartja megoldhatónak. E választás azonban egy súlyos problémát vet föl, jelesül a körkörös magyarázat problémáját – ahogy azt a szerző maga is bevallja (17–18, 80) –, amely az alábbi, az elméletben egyaránt fontos szerepet játszó két állítás miatt áll elő:

1. A szubsztanciákat a trópusnyalábok segítségével azonosítjuk.
2. A nyalábokat alkotó trópusokat mint partikuláris minőségeket az őket hordozó szubsztanciára hivatkozva azonosítjuk.

Ujvári szofisztikált megoldást kínál a körkörös magyarázat elkerülésére: a szubsztanciák konkrét létezők, és mint ilyeneknek kvalitatív vonásaik mellett téridőbeli pozíciójuk is szükséges az azonosításukhoz (91–93). Ezzel szemben a trópusok absztrakt létezők, hiszen egy adott téridő régióban több is jelen lehet belőlük. Az persze igaz, hogy a téridő pozíció partikularizálja a trópusokat – és mivel ezek mindig egy adott téridő régióban léteznek, ezért nevezhetők absztrakt partikuláréknak –, azonban nem individuálja őket. A szubsztanciák és a trópusok tehát eltérő azonosságfeltételekkel rendelkeznek, és az előbbiek individuációjában csak egy részt képviselnek a nyalábot alkotó trópusok, így valójában nincs szó körköröségről. Ez a megoldás azonban azt is jelenti, hogy a könyvben kifejtett trópusnyaláb-elmélet nem nyújt teljes redukív számadást a kontinuáló dolgok mibenlétéről.

Térjünk most át a másik kérdésre, vagyis arra, miért tartja a szerző a trópusnyaláb-elméletet jobbnak riválisainál, a csupasz szubsztrátumot feltételező koncepciónál és az univerzalista nyalábelméletnél! Ami az első vetélytársat illeti, Ujvári egy igen érdekes és szokatlan ellenvetést hoz föl a csupasz szubsztrátumokkal

operáló elmélettel szemben, nevezetesen azt, hogy az elmélet által posztulált tulajdonságok nélküli szubsztrátumok éppen azon feladatok egyikére alkalmatlanok, amelynek betöltésére bevezetésre kerültek, jelesül az individuátori feladatra (31–33). Ujvári szerint ugyanis az elmélet nem zárja ki, hogy két különböző szubsztanciának azonos legyen a szubsztrátuma, márpedig ekkor a két szubsztanciának is azonosnak kéne lennie, ami ellentmond az eredeti föltevésnek. Sőt, Ujvári azt is állítja – és ez talán meggyőzőbb –, hogy két különböző szubsztancia szubsztrátumait fölcserélhetjük, ami által az elmélet szerint az eredeti szubsztanciák maguk is fölcserélődnének. Ám ebből semmit sem vennénk észre, lévén a szubsztrátumok semmilyen észrevehető kvalitatív vonással nem rendelkeznek.

Bevallom, számomra nem túl meggyőzőek ezek az érvek. A probléma az, hogy nem világos – és Ujvári sem teszi világosabbá –, mik lennének maguknak a csupasz szubsztrátumoknak az azonosságfeltételei: hogyan kell értenünk azt, hogy két különböző szubsztancia azonos szubsztrátummal mint alkotórészrel rendelkezik? Ezek talán numerikusan azonosak – de hát az elmélet szerint kettő van belőlük! Vagy kvalitatíve azonosak – de hát a szubsztrátumoknak nincsenek kvalitatív vonásaik! Azt ugyan a szerző is elismeri, hogy a csupasz szubsztrátumok „az azonosságfeltételeik hiányától szenvednek” (34), valamilyen kézzelfogható azonosságfeltétel nélkül viszont nehéz értékelni Ujvári érveit. Ha elfogadjuk a primitív „ezség”, a tovább nem elemezhető partikularitás koncepcióját a szubsztrátumokkal kapcsolatban, akkor Ujvári érvei nem konkluzívak. Márpedig a szubsztrátumoknál éppen ezt szoktuk lényegi vonásuknak tekinteni.<sup>1</sup> A csu-

<sup>1</sup> Lásd például Tőzsér János, *Metafizika*. Budapest, Akadémiai Kiadó. 2009. 69. és 80.



pasz szubsztrátum funkciója tehát pontosan az, hogy mint tovább elemezhetetlen partikuláris elem biztosítsa a partikuláris szubsztanciák azonosságát, függetlenül azok – különböző vagy azonos – kvalitatív vonásaitól. Álláspontom szerint tehát két különböző szubsztancia *per definitionem* nem rendelkezhet ugyanazzal a szubsztrátummal, és nem is cserélhetik föl szubsztrátumaikat.

Az imént bemutatott érvekhez képest Ujvári kissé mellékesen tárgyalja a csupasz szubsztrátum koncepciójával szembeni, talán megszokottabbnak mondható ellenvetést. Arról van szó, hogy önmagában ellentmondásosnak tűnik azt állítani, hogy a csupasz szubsztrátum egyáltalán nem rendelkezik kvalitatív vonásokkal, és így *a fortiori* lényegi tulajdonságokkal sem. Azonban jogosan vetődik föl a kérdés, hogy a „tulajdonságok hordozója”, vagy a „nincsenek magának tulajdonságai” kifejezések nem lényegi vonásokra utalnak-e? (35). Ha viszont ragaszkodunk a lényegi tulajdonságok hiányához, akkor előáll az azonosságfeltételek hiányának már említett problémája.

Ami a másik rivális koncepciót, az univerzalista nyalábelméletet illeti, ezzel szemben a mű különböző részein több különböző ellenvetést hoz föl a szerző (37–52, 124–130).

1. A szubsztanciák individuációját biztosító nyalábokat alkotó univerzálék szükségszerű létezők, míg maguk a szubsztanciák vagy kontinuáló dolgok nyilvánvalóan kontingensen léteznek. Honnan ered e kontingencia?
2. Az univerzalista nyalábelmélet ultraeszcencialista, hiszen az elmélet szerint a nyalábot alkotó összes tulajdonság kivétel nélkül hozzájárul a dolog azonosságfeltételéhez. Így az elmélet nagyon erősen ki van téve az időbeli változás problémájának, mivel bármilyen lényegtelennek vagy aprónak tűnő válto-

zás is megsemmisíti a dolog önazonosságát. Ugyanakkor egy ezzel párhuzamos modális probléma is fellép, hiszen ha minden nyalábbalkotó tulajdonság lényegi, az kizárja, hogy a dolognak kontingens tulajdonságai is legyenek, ez pedig kontraintuitív következmény.

3. A megszorítás nélküli tulajdonságnyaláb-elméletből az következik, hogy a dologra vonatkozó minden tulajdonság-attribúció analitikus állítást eredményez, ami szintén kontraintuitív eredmény.
4. Ujvári szerint a legerősebb ellenérv az univerzalista nyalábelmélettel szemben az, hogy elköteleződik az erős Leibniz-elv mellett. Mivel az elmélet szerint a partikuláris tárgyakat tulajdonságaik együttese individuálja, ezért lehetetlennek kell tartania azt, hogy két numerikusan különböző tárgy pontosan ugyanazzal a tulajdonságokkal rendelkezzen. Márpedig ez intuitíve nagyon is lehetséges, gondoljunk csak az Ujvári által sokat hivatkozott tömegárúk példájára. Az univerzalista nyalábelmélet híve ráadásul nem hivatkozhat arra, hogy csak a gyenge Leibniz-elv mellett vállal elköteleződést, hiszen ez utóbbi a tér-időbeli tulajdonságoknak a nyalábba történő bevonásával egyszerszintű nem-tiszta tulajdonságokat is az individuációért felelős nyaláb részének tekint. Ezek azonban olyan tulajdonságok, melyeket nem tudunk más partikuláris létezőkre való hivatkozás nélkül definiálni. Ezzel a manőverrel kiegészítve az elmélet tehát nem szolgálhatna a partikuláris tárgyak redukzív magyarázatával, márpedig éppen ez a célja.

Mivel saját maga is a nyalábelméletek felé orientálódik, ezért nem meglepő, hogy Ujvári szerint az általa képviselt trópusnyaláb-elmélet legfontosabb előnye éppen

az, hogy nincs kitéve a nyalábelmélet univerzalista változatával szemben fölhozható ellenérveknek. Lássuk, miért!

1. Az első két ellenvetés voltaképpen megválaszolható az univerzalista nyalábelmélet keretei közt is. Az elsőnél például az a helyzet, hogy a szubsztancia kontingens mivoltáról számot adhatunk a nyalábalkotó tulajdonságok együttes előfordulásának kontingens relációjára hivatkozva. A trópusnyaláb-elmélet alapján ráadásul úgy is érvelhetünk, hogy a trópusok nem szükségszerű létezők: időben változnak, keletkeznek és pusztulnak (125–126).
2. A második ellenérvet meg lehet válaszolni az arisztotelészi fajta-univerzálékra hivatkozva, azt állítván, hogy csak az ezek által alkotott nyaláb tartozik hozzá voltaképpen a dolog önzonosságához. Persze bármilyen más alapon is kijelölhetünk a teljes tulajdonsághalmaznál szűkebb, lényegi nyalábot, mely a dolog időn át tartó azonosságát biztosítja, csak félfő, hogy döntésünk önkényes lesz. Ujvári is egy ilyesfajta esszencialista megoldáshoz folyamodik. Ezzel a megoldással ráadásul legalább részben a kontingens predikáció problémáját is megoldja a szerző, hiszen a perifériához tartozó nyalábalkotó trópusok nem vesznek részt a dolog identifikációjában, így azok tulajdonítása nem eredményez tautológiát (141).
3. A trópusnyaláb-elmélet további előnye, hogy itt – szemben az univerzalista nyalábelmélettel – egy kontinuuál dolog azonosságfeltételeiben mindig hivatkozunk valamilyen partikuláris tárgyra – hiszen a trópusokat magukat így azonosítjuk –, ezért elkerülhető, hogy bármely, akár teljesen önkényes tulajdonságegyüttes önálló partikulárét alkosson (124).
4. Mivel az elméletben trópusok, azaz partikulárékként fölfogott tulajdonságok

szerepelnek, ezért az elmélet alapján lehetetlen, hogy két dolog ugyanazokkal a nyalábalkotó tulajdonságokkal rendelkezzen. Így bár ebből az elméletből is következik az erős Leibniz-elv, ez azonban semmiféle hátránnyal nem jár.

Az elmélet ezen előnyei ellenére ugyanakkor Ujvári óva int attól, hogy a trópusokat úgy tekintsük, mint „univerzális gyógyírt a metafizika összes problémájára” (79–80). Nem gondolja például, hogy egyáltalán nem kell univerzálékát is fölvennünk ontológiánkba: a relációkat szerinte csak univerzálékként értelmezhetjük. Ezt a könyv egyik legvilágosabb érvelése során mutatja ki a szerző, ahol is amellet foglalt állást, hogy ahhoz, hogy a trópusnyaláb-elméletben döntő szerepet játszó *együttes előfordulás* relációjáról számot adjunk, az szükséges, hogy e relációt univerzáléknak tekintsük (103–108). Hasonlóan vélekedik Ujvári az úgynevezett meghatározandó monadikus tulajdonságokról (109–113), valamint a változatlan, absztrakt tárgyról is (178). Valójában csak az eseményekre és a szubsztanciákra nézve javasol trópus-elméleti elemzést. Mindkét ontológiai típus esetében az a helyzet ugyanis, hogy a trópusnyaláb-elmélet lehet egy plauzibilis közbülső megoldás az egyaránt szélsőséges partikularizmus és univerzalizmus között.

Már láttuk, hogy könyvében Ujvári valamiféle esszencialista trópusnyaláb-elméletet képvisel, amely nem egy egységes, hanem két különböző ontológiai szinten elhelyezkedő nyalábbal dolgozik. Az így kapott kétszintű nyalábelméletben a „belső”, lényegi trópusokat tartalmazó nyaláb elemeit a Peter Simonstól kölcsönözt, husserli eredetű megalapozási reláció kapcsolja össze (161–173). A megalapozási reláció a relátumként szereplő trópusok kölcsönös feltételezettségét fejezi ki. Például egy kiterjedéstrópus és egy színtrópus kölcsönösen feltételezik egymást, hiszen egyik sem

fordulhat elő a másik nélkül. A konkrét tárgyak lényegét alkotó szűkebb nyalábban szereplő trópusokat tehát nemcsak az együttes előfordulás relációja köti össze, hanem az ennél erősebb megalapozási reláció is. Mivel tehát itt kölcsönös feltételezettségről van szó, ezért e szűkebb nyaláb összetétele nem tetszőleges, hiszen nem áll fenn a szubsztanciát alkotó bármely trópusok közt. Ez biztosítja, hogy valódi, nem pusztán önkényesen kiválasztott lényegi tulajdonságokról van szó. A második, „külső” szinten helyezkedik el a szubsztancia individuációjában részt nem vevő akcidentális tulajdonságok nyalábjá, amelyek esetében az összekötő reláció továbbra is az együttes előfordulás marad, helyet teremtve így az akcenciák kontingenciájának.

A trópusnyaláb-elméletnek ez a konstrukciója tehát a lényegi szűkebb trópusnyaláb bevezetésével elkerüli az ultraeszszencializmus csapdáját anélkül, hogy a csupaszb szubsztrátumot is fölvenné ontológiájába – ami ráadásul nem lenne képes megmagyarázni a kvalitatív változások radikális keletkezést és pusztulást eredményező hatását (160). Mindazonáltal természetesen – ahogy azt a szerző nyilvánvalóvá is teszi (96) – ez a megoldás egy erős értelemben vett egyedi esszencializmus mellett köteleződik el: minden egyedi dolog rendelkezik egy csak rá jellemző tulajdonságpartikulárék által meghatározott lényeggel. Ujvári a kétszintű trópusnyaláb-elmélet segítségével arról is számot ad, hogyan válik absztrakt trópuspartikulárék egy nyalábjá konkrét szubsztanciává (164–165), végezetül pedig az utolsó fejezetben szintén e konstrukcióra támaszkodva megoldást javasol az időbeli azonosság és a változás klasszikus problémáira is. A megoldás lényege abban áll, hogy a szűk nyalábjában található, az egyedi lényegért felelős meghatározott trópusok rövid életűek is lehetnek, megengedve így a változást, míg a fajtalényegét alkotó, meghatározan-

dó trópusok viszonylag állandóak, így biztosítván a szubsztancia időbeli fennállását a változások közepette.

A könyv egyik legizgalmasabb vonása kétségkívül az Ujvári által föl vállalt empirista szellemű metafizikai megközelítés, ami utalásszerűen végig jelen van a műben. Manapság ugyanis alapvetően ellentétesen szoktuk gondolni az empirista szellemiséget és a metafizikai elméletalkotást: míg az első azt vallja, hogy elméleteink végső ítéelőszéke az érzéki tapasztalat, addig az utóbbi arról igyekszik számot adni, hogyan épül föl a világ, függetlenül attól, hogy mi hogyan ismerjük meg azt. Ujvári ezzel szemben mintha azt a fölfogást tenné magáévá, hogy a metafizikai elméleteknek is összhangban kell lenniük azzal, ahogyan számunkra a világ perceptuálisan adódik. Ha jól értem, e követelmény felvállalását hivatott kifejezni a kanti perzisztencia-feltételre mint a *tapasztalat* lehetőségének transzcendentális feltételére történő hivatkozás és támaszkodás.

E fölfogás jelenlétére utal például az, hogy Ujvári a nyaláb- és a szubsztrátumelméletek különbségénél is sokat hivatkozik arra, hogy az első megfelelőbb egy empirista hajlamú filozófus számára, hiszen így a szubsztanciák verifikálható azonosságfeltételeihez jutunk (29, 36–37), sőt egy helyen úgy nyilatkozik, hogy „a nyalábelmélet ismeretelméletileg motivált metafizikai elmélet” (124). A klasszikus brit empiristákat kifejezetten a trópusnyaláb-elmélet előfutárainak tekinti (78–79), akik Ujvári szerint a fizikai tárgyakat érzékelhető kvalitatív vonásaik alapján tartották azonosíthatónak. Maguk e minőségek tekintetében ugyanakkor nominalista álláspontot foglaltak el, vagyis azt vallották, hogy csak a konkrét tárgyak partikuláris tulajdonságai léteznek. Sőt, a kétszintű elméletben szereplő lényegi trópusok kölcsönös megalapozásának eszméjét is az érzéki tapasztalat fenomenológiájából eredezteti

a szerző, hiszen bizonyos absztrakt kvalitatív vonások (alak, mozgás stb.) a percepcióban és a képzeletben fenomenológiailag elválaszthatatlannak tűnnek egymástól és az őket hordozó tárgytól (158–160). Az empirista ismeretelméleti szempont tehát végig erőteljesen jelen van a könyvben, és ez talán egy némileg szisztematikusabb tárgyalást is megengedett volna az elszórt megjegyzéseken túl.

Ezzel elérkeztünk Ujvári könyvének talán egyetlen, ám nem lényegtelen fogyatékoságához: miközben a mű gondolatilag rendkívül alapos, a megfogalmazás módja és a szerkesztettség korántsem olvasóbarát. A szerző nem helyez túl nagy súlyt arra, hogy az olvasó intuícióna apelálva igyekezzen világossá tenni a problémákat és elfogadtatni a megoldásokat. Már a Bevezetés is kissé „hegeli” jellegű, azaz megértése már eleve feltételezi azt az egész problémaköteget és fogalmi apparátust, amelyet aztán magából a könyvből ismerünk majd meg. Itt olyan fogalmakkal és fogalmi megkülönböztetésekkel találkozhatunk, amelyek elsőre furcsának tűnhetnek, ugyanakkor nem kapnak kellő megvilágítást. Például: „partikularizált tulajdonság” vs. „tulajdonságpantikularé” (15), „univerzálé-partikularé” vs. „absztrakt-konkrét” (21). A mű főszereplői, a trópusok is csak az 57–58. oldalon kapnak meghatározást.

További nehézséget okoz az olvasó számára, hogy a tárgyalt elméletek és álláspontok sajnos igencsak szétszórtan, és ebből fakadóan redundáns módon kerülnek kifejtésre. Például az arisztotelészi szubsztancia-fölfogásról egyaránt olvashatunk a 12., a 44., a 96. és a 189. oldalon, azonban sehol sem kapunk egy általános érvényű definíciót. Éppígy a trópusok azonosításának problémája – ami az egyik kulcsprobléma az elméletben – már a Bevezetésben előkerül, a II. fejezet végén ismét fölbukkan, majd a részletes tárgyalásra

a III. fejezetben kerül sor. Ám amikor már azt hinnénk, a kérdést a szerző fölvetette és megválaszolta (93–95), majd más témára tér át, a probléma újból előkerül (101). A talán legkirívóbb esetet a könyv vége felé találjuk, ahol a konstitúcioreláció értelmezésének és a koincidens dolgok lehetőségének, illetve e lehetőségek megszorításának problémái szinte ugyanolyan mondatokkal kerülnek kifejtésre kétszer, hét oldal különbséggel (195 és 202). E tördettségük miatt a gondolatmenetek sajnos néha igencsak nehezen követhetőek, és nem mindig világos, mi a szerepük az argumentáció előrehaladásában.

Ha azonban megbirkózunk a „prezentáció módjának” e nehézségeivel, akkor egy rendkívül alapos, hatalmas gondolati igénynyel megírt ambiciózus műhöz jutunk. Ujvári szigorú következetességgel vet számot álláspontja következményeivel, sehol nem spórolja meg magának a részletes érvelést, az egyes gondolatmenetek összes folyamánának föl kutatását és az azokkal történő szembenézést. Éppen ezért Ujvári könyve egyfajta metafizikai kisenciklopédiának is használható, hiszen a jelzett fő csapásirány mellett foglalkozik még a perdurantizmus és az endurantizmus vitájával (142–147), csakúgy, mint az események ontológiai státuszának kérdésével (66–76), s a Cambridge-változások problémája éppúgy előkerül (191–204), ahogy a radikális keletkezés lehetőségének kérdése (205–210). Mint már a recenzió elején leszögeztem tehát, valóban egyedülálló, kivételes gondolati igényességgel megírt, bátor munkát vehet a kezébe az olvasó, amelyet remélhetőleg sok hasonló követ még a magyarul író filozófusok tollából.<sup>2</sup>

<sup>2</sup> Ezúton szeretném köszönetemet kifejezni Bárány Tibornak, aki a recenzió szövegének korábbi változatához fűzött hasznos javaslataival segített a végleges verzió megírásában.

## Metafizikai hullámvasút

Tőzsér János, *Metafizika*. Budapest, Akadémiai Kiadó. 2009. 262 oldal

Tőzsér János *Metafizika* című könyve igazi szellemi kaland a kortárs analitikus filozófia labirintusában. Tankönyv, még hozzá hézagpótló tankönyv, hiszen a hazai piacon Houranszki Ferenc 2001-ben megjelent *Modern metafizika* című műve óta hasonló ambícióval megírt metafizikai munkával nem lépett színre szerző.<sup>1</sup> Houranszki a metafizikát a teológiához viszonyítva definiálja, így magyarázkodni kényszerül, miért nem tárgyalja Isten létének kérdését, holott az a metafizika klasszikus problémái közé tartozik. Tőzsérnél – miután a természettudományok viszonylatában jelöli ki a diszciplína helyét – ez a kérdés fel sem merül. Houranszki könyve laza szerkesztésű: felsorakoztatja a metafizika hagyományosan elfogadott témáit, problémákat vet fel és tárgyal, de a különféle problémákra adott megoldásjavaslatok közti belső összefüggéseket nem feszegeti. Tőzsér könyve másképp strukturálódik, amennyiben az egyes toposzok lehetséges megoldásai közti összefüggésekre épít. Mindez, hangsúlyozom, nem Houranszki könyvét marasztalja el, ám jelzi, hogy az egyes té-

maköröket markánsabban, belső kapcsolatrendszerük felől is lehetséges tárgyalni, és egyben azt is jelzi, hogy megérett az idő egy, a jelen analitikus (revizionista)<sup>2</sup> metafizika problémáit átfogóan tárgyaló tankönyv kiadására.

A legalapvetőbb kérdés az: vajon tényleg a várva-várt metafizika tankönyvet tartja a kezében az olvasó? A *Metafizika* nem könnyű olvasmány az analitikus metafizikában járatlan olvasó számára. (Ne feledjük: a célközönség az egyetemi hallgató!) Tőzsér nem köt kompromisszumot: egy pillanatra sem hajlandó megszakítani a szigorú logikájú levezetést és magyarázni, szemléltetni mondandóját. Írása ugyanakkor alapmű, mely szerencsésen ötvözi egy detektívregény szövevényességét és izgalmait egy reaktor kapcsolási rajzának aprólékosságával és strukturáltságával. Mint minden valamire való detektívregény, ez az írás is megfeszített figyelmet kíván olvasójától.

<sup>2</sup> A revizionista metafizikát Strawson nyomán (Peter Frederick Strawson, *Individuals: An Essay in Descriptive Metaphysics*, London, Methuen, 1959) a deskriptív metafizikával szokás szembeállítani. A revizionista nem fogadja el, hogy mi, emberek rendelkezniünk a világot leíró alapvető fogalmi szerkezettel, így természetes intuíciónk egy részét – ha egy koherens elmélet megalkotásához ez szükséges – feladja.

<sup>1</sup> Houranszki Ferenc *Modern metafizika* (Budapest, Osiris Kiadó, 2001) című munkáján kívül a *Filozófiai kalauz* (A. C. Grayling szerk. Budapest, Akadémiai Kiadó, 1997) szentel egy fejezetet a metafizika tárgykörének.

A könyv problémaorientált, és mindenfajta történeti megfontolást elvet. (Az analitikus filozófia történetével az első fejezetben egyetlen bekezdés [!] erejéig foglalkozik.) A szerző kétszázhatvankét fesszesen szerkesztett oldalon tárgyalja diszciplínája legfontosabb kérdéseit. Szórványosan alkalmazott, de életszerű példái, a közvetlen hangnem és az analitikus filozófia jellegzetes gondolatkísérletei remekül ellenpontozzák a bonyolult, többszörösen összetett gondolatmeneteket. E filozófiai diszciplína egyik legemblematikusabb ismerve „no nonsense” attitűdje és stílusa. Az olvasás során mindvégig úgy érezhetjük, mintha kíméletlenül érvelő „analitikusokkal” bonyolódtnak volna heves vitába egy hangulatos pesti kiskocsmában.

A szerző a könyv minden fejezetét az adott problémára vonatkozó természetes intuícióink szabatos filozófiai rekonstrukciójával nyitja, majd – a filozófiai kérdés megfogalmazása után – felsorakoztatja a rendelkezésünkre álló bizonyítékokat a lehetséges megoldások mellett és ellen. Mindeközben – mértékkel<sup>3</sup> – tájékoztat minket arról, milyen álláspontot foglalnak el a kérdés nemzetközi szakértői, esetleg ő maga.

Tőzsér *hisz* a kortárs analitikus metafizika erős programjában. Úgy gondolja, a tárgyalt metafizikai kérdések filozófiai eszkö-

zökkel megoldhatók, így a metafizika nem csupán fogalmak tisztázására szolgál (ami mellelleg önmagában sem jelentéktelen cél), s nem is nyugvópont nélküli folytonos kérdésfelvetések és azokból sorjázó válszok folyama (*philosophia perennis*),<sup>4</sup> hanem tényleges, *végleges* válaszok és az ezekből felépíthető *végző, átfogó rendszer* után kutat. Ezért kell világos szerkesztéssel, pártatlan attitűddel és szigorúan felépített taxonómiákkal felvérteznie magát. Ennek fényében talán jobban értjük és értékeljük a szerzőnek azt a kijelentését, miszerint „Egy kézzelfogható eredményt [...] az analitikus filozófia máris felmutathat, és ez példanélküli a filozófia történetében. Nevezetesen azt, hogy napjainkban szinte valamennyi fontosabb metafizikai kérdés vagy probléma kapcsán *előttünk áll a különböző lehetséges elméletek teljes vagy közel teljes taxonómiája.*” (25). Tartsuk szem előtt: ha történetesen igaza van, akkor egy alapos nekigyürkőzés után letehetjük majd a lantot, és felhagyhatunk a metafizika művelésével...

Az ambiciózus program hallatán az első felmerülő gondolat: az episztemológiai aggály. Ha a metafizikában tárgyalt entitásokhoz közvetlenül nem férnénk hozzá, nem tárhatnánk fel például a tárgyak tulajdonságaira, azonosságára, különbségére vagy bármilyen, az elmétől független létezőkre vonatkozó igazságokat. A kanti konceptuális metafizika elutasítása és annak hangsúlyozása, hogy metafizikai vizsgálódásaink közvetlenül az elménktől független világra irányulnak, a szerző által felépített konstrukció sarokköve lehetne. Tőzsér így fogalmaz: „[...] abból, hogy számunkra a világ mindig *valamilyen módon* adódik, *nem következik* az, hogy nem, vagy

<sup>3</sup> Az analitikus hagyományok szellemében Tőzsér a főszövegben szinte egyáltalán nem foglalkozik történeti kitekintésekkel. Ugyanakkor a gazdag irodalomjegyzékből kitűnik, tisztában van az egyes problémák filozófiatörténeti vonatkozásaival, és munkájában szándékosan alkalmazta a jóindulatú rekonstrukciónak az analitikus filozófiában elfogadott elvét: az egyes szerzők gondolatmeneteit történeti kontextusukból kiemelve, érthető, világos formában építi fel, ha azokat újabb érvekkel támogathatja meg, úgy ezt megteszi, majd kimutatja, hogy az érvelés ellentmondásos, vagy ellentmondás-óktól mentes.

<sup>4</sup> Érdemes felhívni a figyelmet arra, hogy Tőzsér itt szembehelyezkedik korábbi, *Jártékok és nyelvjátékok* (Budapest, Kávé Kiadó, 2001) című könyvében kifejtett álláspontjával.

nem közvetlenül adódik. A mód ugyanis, *ahogyan* a valóságra gondolunk, illetve *ahogyan* a valóságot észleljük, *nem közvetítő elem* az elménk és a valóság között” (23). A szerző azonban szándékosan nem halad tovább ezen az úton: könyve metafizika – azaz ontológia –, nem ismeretelmélet. Indokolt az olvasó hiányérzete: a szerző posztulálja a konceptualizmus vereségét, de mivel részletekbe nem bocsátkozik (kétségtelen, itt nem ez a kitűzött cél), a későbbiekben joggal érezhetjük úgy, hogy pusztán jóindulatú előfeltevés a lendületesen megszerkesztett építményt szilárdan megalapozottnak látni.

Miután Heidegger azonos címmel elhangzott híres előadásával felelve saját „Mi a metafizika?” című, bevezető jellegű fejezetben meghatározza munkája tárgyát, s kijelöli helyét az egyes irányzatok között, Tózsér nekilát, hogy a modern analitikus metafizika alapvető kérdéseit módszeresen bemutassa, mégpedig a következő tagolás szerint: tulajdonságok, fizikai tárgyak, fizikai tárgyak létezése az időben, események, lehetséges világok és a test-lélek probléma. Egy recenzióknak nem lehet célja az ismertetett könyv rekonstrukciója; ezért a következőkben csak a második és harmadik fejezetet tárgyalom részletesebben mindenekelőtt azért, hogy bemutassam: az egyre bonyolódó elemzésben az egyes kérdésekre adott válaszlehetőségek milyen indázó következménylancolatot verítenek előre.

Tózsér átgondolt tematika szerint építkezik. Nem véletlenül kezdi könyvét éppen a tulajdonságok problémájának tárgyalásával. Ahogy kiderült: metafizikája ontológia. Arra keres választ, milyen *fajta* dolgok léteznek a világban, így az, hogy a tulajdonságok kérdéskörével kezdi, egyben megalapozza a továbbiakban tárgyalt toposzokat is, hisz a fizikai tárgyak mibenlétéről, perzisztálásáról, változásairól valott nézeteink elválaszthatatlanok attól a

kérdéstől, minek tekintjük a tárgyak tulajdonságait. Miközben áttekintjük a szerző tematikáját, óhatatlanul az az érzésünk támad, „ez nem lehetne másképp”. Ez azonban nem igaz. A szerkesztés Tózsér elkötelezettségét tükrözi egy bizonyos belső logikai rend mellett.

Amennyiben Tózsér könyve ontológia, úgy arra a kérdésre keresi a választ, „milyen dolgok léteznek a világban”. Az erre vonatkozó természetes meggyőződésünket példázza, ha elképzelünk egy partikuláris piros fazekat és egy partikuláris piros almát. Azt gondolnánk, a két partikuláris (individuális) tárgy *numerikusan különbözik*, de legalább egy tulajdonságukban (pirosság) mindazonáltal *azonosak*. Hogy lehetséges ez? A kérdésre adott filozófiai válaszok – az univerzálé probléma – a könyv második fejezetének tárgya. Az univerzálék tárgyalása során a szerző eltér az elméletek hagyományos felosztásától (realizmus, nominalizmus és konceptualizmus), majd – miután e devianciára tisztességgel felhívja figyelmünket – új taxonómiát javasol: a hagyományosan nominalizmusként emlegetett állásponton belül elkülöníti a valódi nominalistákat (szerintük *csak* partikuláris tárgyak léteznek) és a trópuselmélet híveit, akik csupán az univerzálék létezését tagadják, a tulajdonságok (mint partikulárek) létezését azonban megengedik.

A realisták érvelését – akik „a tárgyak típusazonosságát az univerzálék numerikus azonosságára vezetik vissza” (31) – becsületesez szakmunkával összegzi, majd kimerítően tárgyalja az elmélet platóni és arisztotelészi változatát. Az ontológia megalapozása során megmutatja, miért „bizarr” elmélet a realizmus: a tér-idő több pontján egyidejűleg, „egészben és teljesen” megjelenő univerzálék egyszerűen „bizarr” entitások. A „létező dolgok” 2.3.2-ben kibontott tényleges, kimerítő taxonómiája viszont jól példázza azt, hogy a feszített

figyelmet igénylő tematika és az analitikus filozófia erős programjával kapcsolatban kifejtett fenntartásaim ellenére miért remek *tankönyv* Tózsér munkája.

A kérdéses szöveghely alapos olvasása után az alábbi táblázatban vázolt hierarchia bontakozik ki szemünk előtt. Az olvasó, aki specifikus témákat óhajt kutatni,

izgalmas, provokatív kérdésekre bukkan (Döntsük el, miféle partikuláris létező egy árnyék vagy egy gödör!), az viszont, aki csupán egy vizsgára szeretne felkészülni, vagy szeretne szilárd talajt érezni lába alatt, miközben a különféle filozófiai elméletek útvesztőjében bolyong, itt megkapja e fogalmak világos meghatározását.

Ontológia*								
Partikulárék					Univerzálék			
Fizikai tárgyak	Absztrakt tárgyak	Események	Személyek	Lyukak és árnyékok	Monadikus (tulajdonság)		Poladikus (reláció)	
					elsőrendű „pirosság”	magasabbrendű „felesége”	elsőrendű „színes-ség”	magasabbrendű

\*Az ábra a könyvben nem szerepel. A rekonstrukció során az elsőrendű és magasabbrendű univerzálék megkülönböztetését már a 2.3.4. részből vettem.

Térjünk vissza egy pillanatra a fazék és az alma kérdéséhez: vajon az univerzáléként felfogott pirosság a legalkalmasabb arra, hogy e partikuláris tárgyak között fennálló típusazonosság magyarázataként szolgáljon? Nem lenne-e kívánatos – s egyben ontológiai értelemben gazdaságosabb is – megszabadulni a bizonytalan státuszú pirosság-univerzálétól, és megvizsgálni, hátha az (1) e tárgyról megfogalmazott kijelentéseink, esetleg a (2) halmazba tartozás, vagy (3) a hasonlóság jelensége a ludas abban, hogy e két tárgy a pirosság tulajdonságát tekintve azonos? Tózsér a második fejezet 3. részében a nominalizmus három fajtáját tárgyalja: a predikátum nominalizmust, az osztály nominalizmust és a hasonlósági nominalizmust. A predikátum nominalizmust (az elmélet hívei szerint egy piros almában és egy piros kendőben csupán az a közös, hogy mindkettőről igaz módon állíthatjuk a nyilvánvalóan partikuláris „piros” predikátumot) „strucc-nominalizmus”-nak nevezi, miután a valódi kérdést megkerüli, és nem oldja meg a típusazonosság problémáját. A szerző szerint egy megoldástól joggal várhatnánk, hogy megmondja, mi az a világban, ami

egy piros kendőt pirossá tesz, arra azonban nem tér ki, hogy vajon a filozófusok feladata-e, hogy megállapítsák, mi az, ami a pirosságot okozza, vagy esetleg az anyag felépítését vizsgáló fizikusoké?

Ahogy a szerző éles elméjűen kimutatja az osztálynominalizmus tárgyalása során, a „Mi az, ami a kendőt pirossá teszi?” kérdésre akkor se kapunk megnyugtató választ, ha a tulajdonságokat az őket instanciáló tárgyak halmazának (avagy osztályának) tekintjük. Természetesen a koextenzív halmazok problematikája is felkerül az osztálynominalisták bűnlajstromára. A probléma lényege, hogy az osztálynominalista, úgy tűnik, nem lesz képes differenciálni bizonyos tulajdonságok között, amikről józan ésszel könnyű belátni, hogy különbözőek: például a szívvel rendelkezők és a vesével rendelkezők halmaza kontingens módon ugyanazokat az elemeket tartalmazza, holott e két tulajdonság – „szívvel rendelkezni” és „vesével rendelkezni” nyilvánvalóan nem ugyanaz. Tózsér a lehetséges világok tárgyalása során azután még egyszer visszatér e kérdésre.

A harmadik fejezet a fizikai tárgyakkal és azok felépítésével foglalkozik. Már a



kezdet kezdetén ráébredünk, miféle gordiuszi csomó kibogozásának gyürkőzik neki a szerző: minden, amit a tulajdonságokról az előző fejezetben megtudtunk, itt újólág – természetesen más szempontok szerint tagolva – köszön vissza. Tőzsér – miután megismertet minket a népi pszichológiának a fizikai<sup>5</sup> tárgyra vonatkozó öt természetes intuícijával (illetve ezek filozófiai rekonstrukciójával) –, elvarrja a nominalizmus-szálat (számukra a fizikai tárgyak ontológiailag strukturálatlan entitások) – és nekilát kibogozni a trópuselmélet hívei és realisták által felkínált három elméletet: a szubsztrátumelméletet, a nyálábelméletet és a szubsztanciaelméletet. Íme, az összefüggések hálója kissé elnagyolt ábrázolásban:

A szerző két egész bekezdést idéz Locke-tól, mielőtt felírja a szubsztrátumelmélet „képletét”: tárgy = tulajdonságok + szubsztrátum. A 3.2. a szubsztrátum kissé zavaros fogalmát bogozza ki. Tőzsér megfogalmazásával élve: „a szubsztrátum *teoretikus fogalom*, melyet azért kell bevezetnünk, mert nélküle képtelenség magyarázni azt a tényt, hogy bizonyos tulajdonságok szisztematikusan együtt járnak”<sup>6</sup> (67). A szubsztrátum tehát a tárgy tulajdonságait hordozza, „egyben tartja” azokat, de ő maga nem rendelkezik tulajdonságokkal.

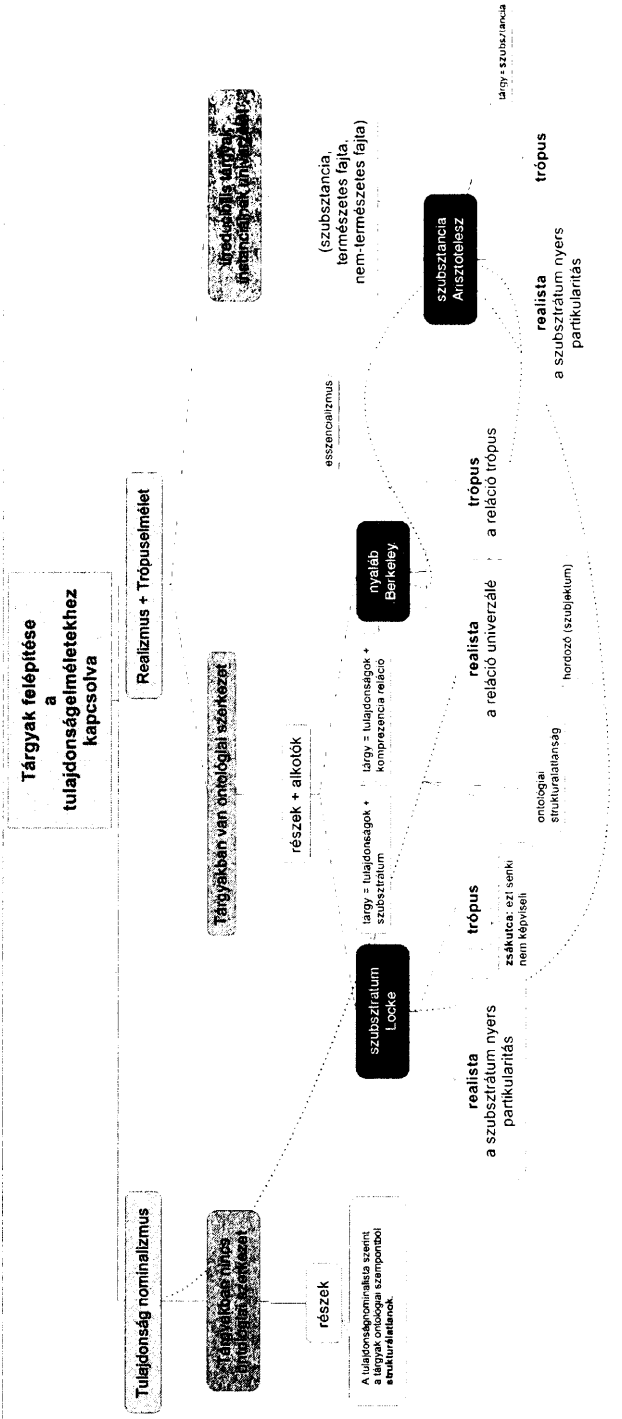
A harmadik fejezet negyedik része szerkezetét tekintve megismétli a megelőző részt: egy Berkeley-től vett idézet után ismerteti a nyálábelmélet lényegét (egy

tárgy nem más, mint tulajdonságainak öszege, azaz nyáláb, valamint a nyálábót összetartó *komprezencia-reláció*), majd megvizsgálja, mit érthet ezen a realista és – ismét csak – a trópuselmélet híve.

A harmadik fejezet ötödik részét a fizikai tárgyra vonatkozó, már tárgyalat elméletekkel szemben – és azok mellett – felhozható érvek pregnáns, kimerítő tárgyalása teszi ki. A nyálábelmélettel szemben három érvet sorakoztat fel. Ezek egyikéről (a változás problémájáról, illetve a változás intuícijához kapcsolható, ezzel szinonim ellenérvről) kimutatja, hogy hatóköre kiterjed a szubsztrátumelméletre is, így ha az érvet elfogadjuk, következményeit is ki kell terjesztenünk a szubsztrátumelméletre is. A második ellenvetés hatóköre a realista- és a trópusnyáláb-elméletet is érinti. Az érv arra kérdez rá, hogy a tárgy egyes tulajdonságai mivel állnak relációban. Térjünk vissza a piros bögréhez egy pillanatra, és tegyük fel, hogy a következőket állíthatjuk róla: (1) piros, (2) bádog és (3) öblös. Az a kérdés, hogy a pirosság tulajdonsága mivel áll relációban, tulajdonképpen arra kérdez rá, vajon *mi az a dolog* (mely tulajdonságok nyálábja), ami piros? A szerző kimutatja, akkor is szákutcába jutunk, ha a pirosság tulajdonságot a tulajdonságok teljes nyálábjához viszonyítjuk, hisz a tulajdonságok teljes nyálábjának megragadása a nyáláb alkotóelemeinek ismerete nélkül nem lehetséges (így a tárgy tulajdonságára vonatkozó mondataink elvesztik informatív erejüket: a „bögre piros” mondatot olyasmiről állítjuk, ami a pirosság, bádogság és öblösség kötege), de akkor is, ha azt állítjuk: az egyes tulajdonságok a tárgy tulajdonságainak csupán egy részével állnak relációs kapcsolatban. (Ez esetben egy látszólag azonos tárgyról szóló kijelentéseim nem ugyanarról a dologról szólnak, hisz e mondatok alanyai tulajdonságok különféle nyálábjai, s ez nyilvánvalóan kontrainuitív.)

<sup>5</sup> Nagyon szerencsésnek tartom, hogy Tőzsér megemlíti az „anyagi” (materiális) és a „fizikai” közti különbséget. A tény, hogy fizikai alatt valami olyasmit értünk, ami egy jövőbeni teljes fizikai elméletben szerepel, a fizikalista megközelítés korlátaira hívja fel a figyelmet.

<sup>6</sup> A megfogalmazás nem szerencsés, hiszen Locke mindvégig ragaszkodik hozzá, hogy a szubsztrátum valóságos létező.



Redukálhatók-e a partikulárek?	Numerikus azonosság-e a típusazonosság?	
	Igen	nem
Igen	Realista nyalábelmélet Realista szubsztátumelmélet	Trópus nyalábelmélet Trópus szubsztátumelmélet
Nem	Realista szubsztanciaelmélet	Nominalizmus

A legkacifántosabb támadás a realista nyalábelmélet (emlékszünk: itt a tulajdonságok univerzálék) ellen irányul: ha lehetséges olyan világ, melyben két numerikusan különböző tárgy valamennyi intrinzikus tulajdonsága azonos (mindkét bögre öblös, bádóg és piros – és más intrinzikus tulajdonsággal nem is rendelkeznek), akkor a Leibniz-elv (megkülönböztethetetlenek azonossága) hamis. Miután azonban a Leibniz-elvet semmilyen körülmények között nem adnánk fel, inkább a realista nyalábelméletről mondunk le, ami így a fedés homályába vész. A realista nyalábelmélet hívei persze azzal védekeznek, hogy megmutatják, ilyen lehetséges világ nem lehetséges. Az általuk megkülönböztetésre használt tulajdonságok azonban (térbeli elhelyezkedés, illetve *haecceitas*, azaz *esszég*) rendre alkalmatlannak bizonyulnak a megkülönböztetendők megkülönböztetésére.

A fejezet hatodik részében tárgyal szubsztanciaelmélet – amit Arisztotelész nevéhez kapcsolhatunk – jelenti a realisták számára az utolsó szalmaszálát. Tőzsér úgy jellemzi ezt a nézetet, mint ami „*a józan középúton jár a szélsőséges nyaláb- és szubsztátumelmélethez viszonyítva*” (88). E kijelentésre sajnos árnyékot vet a következő probléma: miután megismerkedünk a szubsztanciaelmélet három tézisével,<sup>7</sup> kiderül, hogy lehetetlen meggyőző kritériumot találni arra, miért pont az elfogadott

határok mentén szabdalják fel a természetes fajták a világot, és arra se kapunk megnyugtató választ, mi lesz a természetes fajtákba nem sorolható entitásokkal (autókkal, dombokkal, iskolákkal). Ettől függetlenül lelkendezhetünk azon, hogy az elmélet valóban a korábbi két megközelítés legjavát tartja meg, és a problematikus, vagy konstraintívív részeket veti el. Kikerüli a nyalábelmélet *ultraesszencializmusát*, ugyanakkor mégis képes *hordozót* biztosítani a tulajdonságok számára anélkül, hogy ehhez egy szubsztátumhoz hasonló nehézkes fogalmat vezetne be, s teszi ezt úgy, hogy a természetes fajták többszörös instanciálásán keresztül numerikusan különböző partikulárekhoz jut. Sajnos, ha elfogadjuk a szubsztanciaelmélettel kapcsolatban felmerülő alapvető ellenvetéseket, csak két választásunk marad: A korábban tárgyalt megközelítéseket vegyítve vele – Tőzsér szavaival élve – „csúnya hibridet” alkothatunk, vagy az élőlények, elemi részecskék és Isten létezésén kívül kénytelen vagyunk minden mást elutasítani.

A tárgyak mibenlétét firtató „nyomozás” így nyugtalanító hiátust hagy maga után, amiért persze nem feltétlenül a szerzőt, hanem az elméletekkel makacsul szembeszegülő fizikai tárgyakat kell okolnunk. Tőzsér egy pregnáns taxonómiával vigasztal minket: a tárgyak hat statikus<sup>8</sup> elméletét 2 × 2 szempont szerint egy mátrix négy mezőjébe utalja (a szempontok: „Numerikus azo-

<sup>7</sup> Első: a fizikai tárgyak természetes fajta-univerzálék instanciái. Második: A fizikai tárgyaknak az univerzálék nem alkotóelemeik. Harmadik: A fizikai tárgyak ontológiailag alapvető entitások (6.1.).

<sup>8</sup> A tárgyra vonatkozó statikus elméletek – a dinamikus elméletekkel szemben – a tárgyak időbeni változását, perzisztálását nem tárgyalják.

nosság-e a típusazonosság?” és „Redukálhatók-e a partikulárek?”), végül az olvasóra bízva, hogy választ találjon a kérdésre.

Az első három fejezet részletesebb tárgyalása érzésem szerint jól illusztrálta Tőzsér tematikájának erényeit és a vele kapcsolatos nehézségeket. A „Fizikai tárgyak létezése az időben”, „Események”, „Lehetséges világok” és „A test–lélek probléma két aspektusa: a mentális okozás és a fenomenális tudatosság” fejezetek tárgyalása során már csak elnagyolt rekonstrukcióra törekszem, és főleg az egyes toposzoknak a korábbi témákhoz való viszonyát emelem ki, mivel a szerkesztésnek ez a módja (*interconnectivity*) a könyv talán legkomolyabb erényének tekinthető.

A „Fizikai tárgyak létezése az időben” című fejezetben az eddig *statikus* tárgyak mozgásba lendülnek és megelevenednek. A fejezet a következő intuíció-csokorra épít: a fizikai tárgyak időben határolt, létezésük minden pillanatában teljes egészükben prezentálódó, és változásra képes entitások. E természetes intuícióink nem konzisztensek: ha egy tárgy akkor, és csak akkor azonos önmagával, ha minden tulajdonsága azonos, akkor a tárgyak változásra képtelen entitások. Az intuíció-csokor egyik vagy másik elemének feladására kényszerülünk. Az intuíciókból kibontható konfliktus feloldására bevezetjük az *endurantizmus* (hétköznapi vélekedéseinknek jobbára megfelelő elmélet, miszerint egy tárgy minden pillanatban teljes egészében jelen van) és a *perdurantizmus* (a tárgyakat az idő egymást követő pillanataiban reálisan létező részek, „szeletek” aggregátumának tekintő elmélet)<sup>9</sup> megkülönböztetését,

<sup>9</sup> Azaz: ha kék automat pirosra festem, úgy a változás tényét a perdurantista úgy magyarázza, hogy az „autó summa” tárgy „autó T1” nevű része kék és „autó T2” nevű része piros. E reálisan létező részek aggregátuma alkotja a tárgyakat, amelyek így nemcsak térbeli, de időbeli dimenzióval is rendelkeznek.

majd vizsgáljuk az ebből fakadó szerzteágazó metafizikai következményeket.

A *perdurantizmus*sal kapcsolatos érdekesebb témák közé tartozik annak lehetősége, hogy egy tárgy temporális részei *átfedhetik* egymást, vagy az a kérdés, hogy a temporális részeket felszabdalva temporális atomokhoz jutunk-e?

Az ötödik fejezetet Tőzsér az események vizsgálatának szenteli. Vegyük észre, továbbra is tapasztalataink világának *belakása* a cél: a tulajdonságoktól eljutottunk az őket megjelenítő mozdulatlan tárgyakig, majd ezeket mozgásra bírtuk az időben, hogy most azt lássuk, e mozgó világban *események* vesznek minket körül. Szándékoltnan olyan elméleteket tárgyal, melyek elkötelezik magukat az események mint ontológiailag önálló entitások mellett. A józan ész egyszerű belátásain túl filozófiai érvert kovácsolhatunk az események létezésével azok oksági kapcsoltóságából, a változás jelenségéből, vagy akár szemantikai alapon is. (Ez utóbbi lényege az, hogy megmutatjuk, egy eseményt tartalmazó mondat érvényességét az esemény feltételezése nélkül, pusztán az elsőrendű logika alapján nem tudjuk magyarázni.)

Az előző fejezetek metodikája eddigre beérret: az olvasó szinte elvárja, hogy a szerző megvizsgálja, partikuláris vagy univerzális, absztrakt vagy konkrét entitások-e az események.

Tőzsér nem véletlenül hagyta könyve legvégére a „Lehetséges világok” és „A test–lélek probléma két aspektusa: a mentális okozás és a fenomenális tudatosság” című fejezeteket. Mindkét fejezet bizonyos értelemben kakukktójságnak tekinthető. Nézzük először a lehetséges világokat: korábbi témái azt feszegették, „mi van”. Most arra kérdez rá, mi lehetne. Még pontosabban arra, aminek léteznie kell, és arra, ami lehetséges, hogy létezzen. Bár ebben a fejezetben is konkrét metafizikai problémákkal szembesülünk, mégis –

a fejezet rövidségéből is – azt sejtjük, itt inkább egy fontos segédeszköz használatának mikéntjére tanítanak meg minket. A tartalomjegyzéket „lapozgatva” talán úgy gondoltuk, Tőzsér nem tárgyal néhány olyan témát, amit egy hagyományos metafizika tankönyvben aligha szokás megkezdni. Bár ezek a kérdések nem érdemelték ki, hogy fejezeteim alá gyűjtsük őket, itt mégis módunk nyílik majd arra, hogy megismerkedjünk velük: Megtudjuk, a lehetséges világokra támaszkodva, *de dicto* és *de re* kontextusban hogyan értelmezhetjük a szükségszerűséget, a lehetőségességet vagy a kontingenciát.<sup>10</sup>

Persze az ontológiai vezérfonalat visszacsempészi a szerző, hogy kerekké, tematikussá tegye könyvét: a fejezet hátralévő része azt vizsgálja, hogy e lehetséges világok – amelyek annak eldöntésében segítenek minket, mi lehetne, és minek kell lennie – maguk is vitatható ontológiai státusszal rendelkeznek. Milyen értelemben *vannak* a lehetséges világok? – kérdezi Tőzsér. A realisták két elmélettel állnak elő, míg az antirealisták számára e világok léte pusztán fikció.

A könyv legutolsó fejezete a test–lélek probléma két aspektusát vizsgálja: a mentális okozást (mentális állapotainknak, szándékainknak, vágyainknak stb. következményei gyakran cselekedetek, azaz fizikai események) és a fenomenális tudatosságot, azaz azt a jelenséget, hogy élünk, egyes szám első személyű perspektívával rendelkezünk. E fejezet olvasása során erős a gyanúnk: szemben az első öt fejezettel (és részben az imént górcső alá

vett hatodikkal), a test–lélek probléma létjogosultsága nem a téma belső logikájából fakad. A tárgykör kifejtése viszont – annak autonómiája ellenére, vagy pont ennek köszönhetően – alapos, logikus, kimerítő és szigorúan szerkesztett. A szerző végkövetkeztetése szerint ellentétes tendencia rajzolódik ki, ha a test–lélek problémát a mentális okozás felől közelítjük meg – ez esetben akkor kapjuk a legnagyobb magyarázó erővel bíró elméletet, ha a relációt a lehető legszorosabbra fűzzük, például azonosnak tekintjük a kettőt –, illetve akkor, ha a fenomenális tudatosság jelenségének magyarázata a cél: ez esetben a két szféra bizonyos fokú autonómiájának feltételezése mellett kapunk plauzibilis elméleteket. A szerző a végkövetkeztetéséhez a szubsztancia dualista elméletek kritikája után a különféle fizikalista elméletek kimerítő elemzésén, valamint a fenomenális tudatosság jelenségével összefüggésben használatos három érvrendszer (magyarázati szakadék, tudásérv, elgondolhatósági érv) tárgyalásán keresztül jut el.

Örömteli a bőséges felhasznált irodalom mellett<sup>11</sup> az ajánlott irodalom külön kigyűjtése, s az egyébként rendkívül tömény és bonyolult terminológiájú szöveg feldolgozását segítik a fejezet végi összefoglalók és a gondosan összeállított név- és tárgymutatók.

És most tegyük fel újra a kérdést: beváltja-e a szerző ígéreteit? A könyv tematikája, szerkesztése világos, szakmai szempontból kimerítő, nyelvezete sallangoktól mentes és szabatos. Olvasása viszont – jórészt szigorú, egymásra épülő szerkesztésnek köszönhetően – megfeszített figyelmet kö-

<sup>10</sup> Egy esemény – a lehetséges világok terminológiáját segítségül hívva – szükségszerű, ha minden világban bekövetkezik, lehetséges, ha van olyan világ, amelyben bekövetkezik és kontingens, ha aktuális világunkban bekövetkezik, de nem minden lehetséges világban következik be.

<sup>11</sup> A könyvben a felhasznált irodalmak között 41 olyan munka szerepel, amely 2000 óta született.

vetel, így nem vagyok meggyőződve róla, hogy a *Metafizika* az egyetemi hallgatók kedvence lesz. Egy biztos: a szerző a *létező dolgok* elméleteinek aprólékos és sokéves kritikai munka árán rendszerbe foglalt hálóját rajzolja fel, s bár néha meglep minket magabiztossága, miközben végső megol-

dásokat sejtet, s való igaz, a szálak itt-ott elpattannak, a szöveg hol sűrűbb, hol likacsos, de elemzései precízek és korszerűek, nincs elvarratlan szál, így véleményem szerint a *Metafizika* egyszerű, dísztelen, de stabil alapköve lehet a kortárs analitikus filozófia hazai oktatásának.

## Angol nyelvű összefoglalók

### Husserl's Methodological Transcendentalism

LÁSZLÓ TENGYELI

Husserl's 'transcendental idealism' has been a stumbling-block in the eyes of several thinkers, who have been attracted by the phenomenological method. Among these thinkers, not only Ingarden, Patočka or Merleau-Ponty can be mentioned but also Heidegger, who severely criticized the subjectivism and idealism he had detected in Husserl's *Ideas I*. However, it is almost immediately after the publication of *Ideas I* that, in the draft of an amended second edition of the *Sixth Logical Investigation*, published now in the volume XX of the series *Husserliana*, the meaning of Husserl's transcendental idealism begins to change. It is documented by some research texts, now available in the volume XXXVI of the series *Husserliana*, how, between 1914 and 1917, Husserl increasingly realizes that, in *Ideas I*, he did not properly take into account the embodiment of the subject and the fact of intersubjectivity. As a result of this reflection, transcendental idealism takes an entirely new form. Assuredly, Husserl continues to consider the world as constituted by consciousness. That is why he rejects dogmatic materialism or naturalism and insists on what he designates as 'transcendental idealism'. However, he does not shrink from assuming a phase of purely material nature in the history of the world. This fact shows that, in reality, *phenomenological transcendentalism does not entail any commitment to idealism in the customary sense of this word*. Indeed, Husserl envisages a world *without* subjects that actually experience it, adding that such a world is only conceivable as the past of a world *with* such subjects. What is meant here is a past constituted retroactively ('backwards') by these subjects. Since the transcendentalism of backward constitution does not seem to entail any commitment to idealism in the customary sense of this word, it may be misleading to designate it as 'transcendental idealism'. Therefore, it is suggested in the paper that Husserl's phenomenological approach should be characterized as a *methodological transcendentalism*, even if this term is not used by Husserl himself. It is added that only a methodological transcendentalism can find a way out of the irresolvable antinomy between a dogmatic materialism or naturalism and an equally dogmatic idealism.

## Philosophical Naturalism and the Problem of Natural Kinds

FERENC HUORANSZKI

The paper argues that all properties that are essential for a particular or stuff to belong to a certain natural kind are dispositional. Both phenomenal properties (e.g. having a certain color) and macro-structural states (e.g. having certain number of legs) are kind-specific essential properties to the extent that they are the manifestations of the particular's or the stuff's dispositions. The usual arguments against the view that phenomenal and macro-structural properties can be metaphysically essential for kind-membership fail to distinguish properly between the conditions of having a disposition and the conditions of its manifestation. Finally, the paper shows that constitution and genetic makeup is metaphysically essential for kind-membership only if they are necessary for the explanation of the relevant set of dispositional properties. Thus, if 'Twin-Earth scenarios are metaphysically possible – which is not obvious – then they prove that the type of constitution is not essential for kind-membership.

## The Origin of our Species

ÁKOS SIVADÓ

In this paper I study the connection between our social and biological kinds and the way we divide our natural and social environment according to our scientific interests. The article tries to examine the nature of our biological and social kind terms based on the differences in their constitution – and the consequences that follow for the scientific generalizations, explanations and predictions they make possible. I also try to show that these differences render a reductionist methodology untenable.

## Naturalism, Skepticism and Rationality

The Evolutionary Skeptical Argument of Alvin Plantinga

MIKLÓS SZALAI

According to Alvin Plantinga's famous „Evolutionary Argument Against Naturalism” (EAAN), the evolutionary-naturalistic account of our cognitive powers should make us skeptical about their reliability – because the truth, justifiability, implication and other logical-epistemological properties of our beliefs are „invisible” to natural selection, which selects the most adaptive, and not the most truth-conducive beliefs and cognitive mechanisms (or their bearers). The author analyzes Plantinga's argument from an epistemological standpoint. To some extent EAAN is self-refuting, and externalist and common-sense philosophical strategies also may be working against it. However, the conclusion of the article is, that the really effective response to Plantinga's challenge is a non-naturalistic, apriori account (and defense) of the reliability of our cognitive powers.

## Naturalism in Psychology

CSABA PLÉH

The paper first shows that naturalism as such does not clearly show up in present day psychology. It certainly is evidenced, however, in discussing materialistic reductionism. The author goes on to argue that present day naturalism has a better fate than its predecessors. There are two main reasons for this. First, we are approaching themind from two

biological perspectives, both Darwinian and neuroscience models are used to interpret the mind. Second, present day biology with its emphasis on neuronal plasticity and epigenetic models in unfolding the genetic envelop allows the psychologist to entertain naturalism while at the same time avoiding any strict determinism.

### Philosophical Naturalism and the Heterogeneity of the Sciences

GÁBOR ZEMPLÉN – GÁBOR KUTROVÁ CZ

The paper argues that philosophical naturalism, in the light of a view of science emerging as a consensus-view in the social and cultural studies of science, is facing some fundamental problems. On the one hand, the profound and manifold heterogeneity of scientific practice poses serious difficulties to many ontological versions of naturalism – adding arguments to the classical criticism given by Crane and Mellor. On the other hand, methodological interpretations of naturalism seem either inconsistent when relying on a Utopian vision of the unified scientific enterprise, or empty when including itself among the plethora of scientific practices. The paper urges the reconsideration of essential findings of contemporary empirical investigations of science when carrying out philosophical analyses, especially when these are intended to be naturalistic.

---

*A Magyar Filozófiai Szemle szerkesztősége május 31-ig vár magyar nyelvű írásokat a folyóirat 2010/2. számába, körülbelül egyívnyi (40 000 karakteres) terjedelemben. A beérkezett kéziratokat független bírálók értékelik. E szám témája a test.*

A test a nyugati filozófiának valamilyen, gyakran változó értelemben mindig is alapproblémája volt a test-lélek viszony klasszikus kérdésétől kezdve a szónak a nem emberi testekre való átviteléig. A test gyakran szerepelt különböző kozmológiai elméletek kulcsszavaként vagy a politikafilozófia nélkülözhetetlen modelljeként. A szónak ezeket a használatait természetesen rutinszerűen vádolták azzal, hogy félrevezető metaforák foglyaivá teszik a(z égi, politikai vagy más) testekkel operáló kozmológiai, politikai és egyéb elméleteket, ugyanakkor nemigen tudjuk elképzelni ezeket az elméleteket hasonló metaforák használata nélkül. Az emberi test azonban akár a lélekkel való viszonyától és a szó metaforikus használatától eltekintve is a filozófiai vizsgálódás tárgyává válhat; gondoljunk csak a napjainkban népszerű test-fenomenológiai vizsgálódásokra, a saját test észlelésének ismeretelméleti problémáira, az emberi testtel kapcsolatos feminista megközelítésekre, vagy akár a bioetika meggondolásaira. A folyóiratszám semmilyen értelemben nem kíván irányzatos lenni: a főt érintett minden problémával kapcsolatban fölmerülő bármely lehetséges filozófiai álláspont mellett érvelő írásokat várunk a szerkesztőség elektronikus postacímére.

A 2010/3. szám témája a *kauzalitás*. A témához kapcsolódó írásokat szeptember 15-ig várjuk a fenti címen. Az ezzel kapcsolatos felhívást és a lapra vonatkozó további információkat a következő nyomtatott szám megjelenéséig a MaFLa levelezőlistán tesszük közzé (jelentkezés: <http://phil.elte.hu/cgi-bin/mailman/listinfo/mafla>).



## A LEADOTT KÉZIRATOK FORMAI KÖVETELMÉNYEI

1. A LEADÁSRA VONATKOZÓ ADATOK: A tanulmányokat két példányban kérjük elektronikus úton eljuttatni a Szemle szerkesztőségéhez a < filszemle@webmail.phil-inst.hu > címre. Kérjük, hogy az egyik példány a közlésre szánt állapotot tükrözze, míg a másik – az anonimitást biztosítandó – szerzői név és felismerhető önhivatkozások nélkül, abban a formában álljon, ahogyan referálás céljából továbbíthatjuk. A nem tanulmány jellegű kéziratokat (például recenziókat) elegendő a közlésre szánt állapotban leadni.

2. A FŐSZÖVEGRE VONATKOZÓ ADATOK: A kéziratot 12-es betűméretben, másfeles sortávolsággal, Times New Roman betűtípusban, „rtf” formátumban kérjük leadni csak a legszükségesebb formázásokkal (kurzív stb.). Körülbelül egyívnyi (40 000 leütés) terjedelmű szövegeket várunk. A fő szövegben előforduló műcímek kurzivan, a hosszabb (4-5 sornyi vagy annál terjedelmesebb) idézetek külön sorban kiemelve, behúzva, idézőjel nélkül szerepeljenek.

3. ÉVSZÁMKIEMELŐ HIVATKOZÁSI RENDSZER: Tanulmányok esetében a hivatkozott szakirodalmat lehetőség szerint a szövegbe illetve kérjük feltüntetni a következő módon: (Kuhn 1999a. 99.), vagy lábjegyzetben: Putnam 1975b. 237. Szükség esetén megadható az eredeti megjelenés dátuma (Descartes 1644/1996.) és a kötetszám is (Descartes 8/2:72). Használhatók a széles körben elterjedt rövidítések (AT VIII-2. 72), és – különösen klasszikus szövegek esetében – a bevett utalási eljárások (Platón, *Phaidrosz* XXIV. 245c–d, Mt 9,13). A rövidítések feloldását kérjük az irodalomjegyzékben megadni.

4. IRODALOMJEGYZÉK TANULMÁNYOK ESETÉBEN: Az irodalomjegyzéket a következő példának megfelelő formában kérjük összeállítani:

ÖNÁLLÓ KÖTET:

Locke, John 1690/2003. *Értekezés az emberi értelemről*. Ford. Vassányi Miklós – Csordás Dávid. Budapest, Osiris.

GYŰJTEMÉNYES KÖTET:

Henke, Christopher R. – Thomas F. Gieryn 2008. Sites of Scientific Practice: The Enduring Importance of Place. In Edward J. Hackett – Olga Amsterdamska – Michael Lynch – Judy Wojcman (szerk.) *The Handbook of Science and Technology Studies*. Cambridge, Mass., The MIT Press. 354–376.

PERIODIKUMBAN MEGJELENT CIKK, TANULMÁNY:

Schnädelbach, Herbert 1993. A filozófia és az embertudományok. Ford. Mezei György Iván. *Magyar Filozófiai Szemle*, 37/2. 1015–1031.

Általános elvek: A neveket csak az abcérend miatt releváns helyeken fordítjuk meg. A kötet vagy a folyóirat főcímét kurzivan szedjük. Minden értelemegységet ponttal különítünk el egymástól, kivéve a) külföldi szerző esetében a név egyes elemeit, b) a megjelenés helyét és a kiadót (ez utóbbi esetekben vesszőt használunk), valamint c) szóközzel és hosszú kötőjellel kapcsoljuk össze az egymás mellett álló személyneveket, valamint a helységneveket (London – New York). Pont után nagy betűvel kezdünk, az „In” után nem teszünk írásjelet. A „szerk.” rövidítés zárójelbe kerül: (szerk.)

5. EGYSZERŰSÍTETT HIVATKOZÁSI RENDSZER: (Elsősorban *nem tanulmány jellegű* kéziratok esetében használandó, de indokolt esetben alkalmazható tanulmányokban is.) Első előfordulásakor kérjük kiírni a teljes információt, szerzőt, művet, fordítás esetén a fordítót, a megjelenés helyét, idejét, a kiadót és az oldalszámot. Ehhez a 4. pontban lefektetett elvek adnak útmutatást azzal a különbséggel, hogy itt a megjelenés évszáma a kiadó neve és az oldalszámok (periodikumok esetében a periodikum címe és évfolyama) közé kerül. A szerző nevét nem fordítjuk meg, hanem vesszővel választjuk el a mű címétől.

Jan Golinski, *Making Natural Knowledge: Constructivism and the History of Science*, *Cambridge History of Science*. Cambridge, Cambridge University Press. 1998. 37.

Ezek után a főszövegben vagy a lábjegyzetben már elegendő a szerző családnevét, továbbá a megjelenés évszámát feltüntetni, természetesen az adott oldalszámmal együtt. Használható az „uo.” rövidítés is.

EGYÉB Kérjük, hogy a szerző írjon magáról egy rövid (1-3 soros) bemutatkozó szöveget (munkahely, tudományos státus, fő kutatási, érdeklődési terület, fontosabb publikációk, e-mail elérhetőség, ha meg kívánja adni). Kérjük továbbá, hogy a leadott tanulmányokhoz egy rövid angol nyelvű összefoglalót sziveskedjen mellékelni.

# MAGYAR FILOZÓFIAI SZEMLE

2010/1 (54. évfolyam)

A naturalizmus fogalmát többféle értelemben használják a kortárs filozófiában, ezek azonban összefüggő problémakört jelölnek ki. A módszertani naturalizmusi szerint a tudományok episztemikusan felsőbbrendűek minden más megismerő módszernél. A metafizikai naturalizmus szerint az ontológia szempontjából is a tudományoké a végső szó: a legjobb tudományos elméletek mondják meg, hogy mi létezik. Persze, ha ezt elfogadjuk, felmerül a kérdés: mit tesz hozzá a filozófia a tudományokhoz? A naturalisták mondhatják: egyfelől a tudományok és a filozófia folytonosak, a filozófiai érvelés és a fogalomelemzés nem választható el élesen az empirikus kutatásoktól. Másfelől a filozófia olyan kérdéseket próbál megválaszolni, amelyekkel a tudományok maguk nemigen foglalkoznak, mint például a tudomány megkülönböztetése más szellemi formáktól, azaz a demarkáció kérdése, a különböző tudományos elméletek által feltételezett létezők viszonya, a redukció problémái, vagy a természeti fajták kérdésköre. Persze a naturalizmus helytállósága vagy akár az értelmessége is kétségbe vonható: a naturalizmus témája köré felépíthető egy történeti narratíva is a természettudományos és a szellemtudományi hagyomány évszázados vitájáról, a „tudományos filozófia” múlt század végi rivális értelmezéseitől a mai „tudományháborúig”, az anti-pszichologizmustól a naturalizált episztemológiáig és matematikafilozófiáig. Számunk tanulmányai e témákból nyújtanak ízelítőt.

ISSN 0025-0090



9 770025 009005

1200 Ft

**GONDOLAT**

**nka**  
Nemzeti Kulturális Alap

P25591

Pn



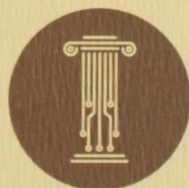
2011 JAN. 27

A Magyar Tudományos Akadémia  
Filozófiai Bizottságának folyóirata

# MAGYAR FILOZÓFIAI SZEMLE

2010/2 (54. évfolyam)

A test



## SZERKESZTŐSEG

Schmal Dániel felelős szerkesztő, Pázmány Péter Katolikus Egyetem  
Ambrus Gergely, Miskolci Egyetem  
Mekis Péter, Eötvös Loránd Tudományegyetem  
Mester Béla, MTA Filozófiai Kutatóintézet  
Ullmann Tamás, Eötvös Loránd Tudományegyetem  
Valastyán Tamás, Debreceni Egyetem

## SZERKESZTŐBIZOTTSÁG

Forrai Gábor, a szerkesztőbizottság elnöke, Miskolci Egyetem  
Almási Miklós, az MTA rendes tagja, Eötvös Loránd Tudományegyetem  
Altrichter Ferenc, University of North Caroline at Wilmington  
Csejtei Dezső, Szegedi Tudományegyetem  
Fehér Márta, Budapesti Műszaki és Gazdaságtudományi Egyetem  
Heller Ágnes, az MTA rendes tagja, Eötvös Loránd Tudományegyetem, The New School for Social Research (New York)  
Hévizi Ottó, MTA Filozófiai Kutatóintézet Lukács Archivum  
Hont István, King's College (Cambridge)  
Kelemen János, az MTA levelező tagja, Eötvös Loránd Tudományegyetem  
Klima Gyula, Fordham University (New York)  
Máté András, Eötvös Loránd Tudományegyetem  
Rózsa Erzsébet, Debreceni Egyetem  
Szabó Zoltán, Yale University  
Tengelyi László, Bergische Universität (Wuppertal)  
Weiss János, Pécsi Tudományegyetem

A szerkesztőség elektronikus postacíme:  
filszemle@webmail.phil-inst.hu  
schmal.daniel@chello.hu

ISSN 0025-0090

Kiadja a Gondolat Kiadó  
Felelős kiadó Bácskai István  
1088 Budapest, Szentkirályi u. 16.  
Telefon: +361-486-1527  
www.gondolatkiado.hu  
<http://gondolatkiado.blog.hu>

Tördelő: Lipót Éva  
Olvasószerkesztő: Böröczki Tamás  
Lapterv: Andorka Tímea és Pintér László  
Nyomás és kötés: *mondAt Kft.* • Budapest • [www.mondat.hu](http://www.mondat.hu)

A lap megjelenését a Nemzeti Kulturális Alap  
és a Magyar Tudományos Akadémia Filozófiai Bizottsága támogatta



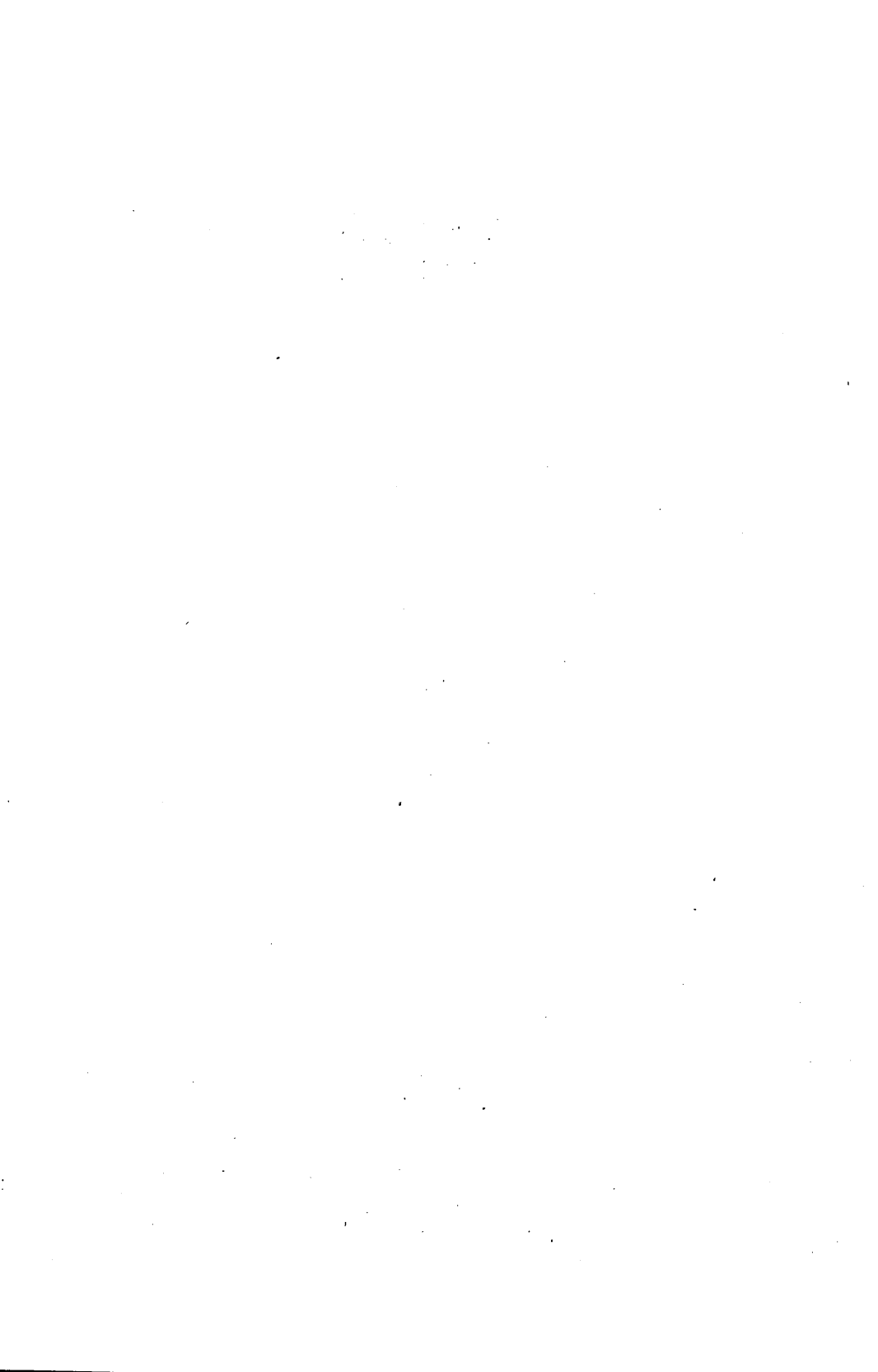
2011 JAN. 27

# MAGYAR FILOZÓFIAI SZEMLE

2010/2 (54. évfolyam)

A Magyar Tudományos Akadémia  
Filozófiai Bizottságának folyóirata

A test



# Tartalom

## TANULMÁNYOK

A test (Mester Béla)	5
SUTYÁK TIBOR: A test–test probléma	10
KICSÁK LÓRÁNT: Transzcendentalitás és testiség a husserli fenomenológiában	32
TAKÁCS ÁDÁM: „Szubjektív tárgyiség”. Husserl fenomenológiai elmélete a test megalapozó szerepéről	51
JOÓ MÁRIA: A feminista elmélet és a (női) test	64
MOLDVAY TAMÁS: A spinozai test mint probléma	81
BALOGH LÁSZLÓ LEVENTE: A Leviatán anatómiája	101
GULYÁS PÉTER: A szimpátia és a belénkszületettség testi értelmezése Descartes-nál	113

## FÓRUM

ESZES BOLDIZSÁR – TŐZSÉR JÁNOS: Analitikus filozófia és fenomenológia	133
---	-----

## SZEMLE

TÓTH ZITAVÉRONIKA: Borbély Gábor és a középkori filozófia	157
---	-----

E SZÁMUNK SZERZŐI	161
-------------------	-----

SUMMARIES	161
-----------	-----





## A test

A *test* valamilyen értelemben mindig is a filozófia egyik alapproblémája volt a test–lélek viszony klasszikus kérdésétől kezdve a szónak a nem emberi testekre való átviteléig, a mértani testektől a közösség testéig. Gyakran szerepelt a test különböző kozmológiai modellek kulcsszavaként vagy a politikafilozófia nélkülözhetetlen modelljeként is. A szónak ezeket a használatait természetesen rutinszerűen vádolták azzal, hogy félrevezető metaforák foglyaivá teszik a (mértani, égi, politikai, állati vagy emberi) testekkel operáló kozmológiai, politikai és egyéb elméleteket, ugyanakkor nemigen tudjuk elképzelni, hogy miképpen lehetne hasonló elméleteket ilyesféle metaforák használata nélkül megalkotni. Természetesen az emberi test az elmével (avagy a lélekkel) való viszonyától és a „test” szó metaforikus használatától eltekintve is a filozófiai vizsgálódás tárgyává válhat, gondoljunk akár napjaink testfenomenológiai vizsgálódásaira, akár a bioetikának a filozófián belül is egyre gyakrabban tárgyalt meggondolásaira. A test így a filozófiában egyaránt megjelenik történeti kérdésként és a kortárs diszkusszió problémájaként: aligha képzelhető el a filozófiatörténetnek olyan korszaka, illetve a kortárs diszkusszióknak olyan irányzata, amelynek ne lenne karakteres álláspontja a testről.

A tanulmányok élén Sutyák Tibor *A test–test probléma* című írása áll, melyben a szerző először arra vállalkozik, hogy áttekintse a test problémájának tárgyalását mind a fenomenológia, mind az analitikus filozófia közelmúltjában. Sorra veszi a leggyakrabban tárgyalt toposzokat (kvália, mentális okozás, mentális és fizikai viszonya) és elméleti példákat (legrészletesebben a *zombikkal* kapcsolatban megfogalmazott érvekkel foglalkozik). Sutyák mind a fenomenológiai, mind az analitikus tradíció teljesítményével szemben kritikus. Mint tanulmánya elején írja, „az egyik oldalon egy masszív rejtély megszerkesztését, a másik oldalon egy klasszikus kérdés korántsem megnyugtató elnyelését tapasztaljuk”. Merleau-Pontyra támaszkodó saját, a *testi elmére* vonatkozó álláspontja kifejtésekor megszívelendő módon óvakodik a(z) egyébként gyakran tapasztalható retorikai túlzásoktól. „[E] kell tekintenünk az elmélet kifejtésében hangot kapó mozgalmár, túlfűtött, időnként prófétikus retorikától, [...] hogy minek az ideje járt

le, mivel kell holnaptól felhagyni, és hogyan változik meg rövidesen úgyszólván minden a reflexióban” – írja, hiszen ez a szóhasználat vagy visszariasztja, vagy félrevezeti az olvasót.

A következő két tanulmány, Kicsák Lóránt és Takács Ádám írása egyaránt a husserli fenomenológia testelméletének bemutatására vállalkozik. Kicsák Lóránt elemzése (*Transzcendentalitás és testiség a husserli fenomenológiában*) olyan alapvető problémák husserli kezeléséből indul ki, mint az esetlegesség és szükségyszerűség, a biológiai evolucionizmus és a világ és öntudat kifejlődéseként leírt fejlődésfogalmak különbsége és az ebből eredő teleológiai problémák. A továbbiakban a szerző a testiségnek a husserli gondolkodás kialakulásában betöltött szerepét tekinti át, megállapítva, hogy az egyik alapvető husserli fogalom, „a transzcendentális szubjektivitás fogalma (mondhatnánk fenomenális tartalma) épp a testiség elemzésével bővült ki jelentős mértékben (tett szert egyáltalán valamilyen tartalomra) [...] a transzcendentális szubjektum a descartesi típusú szubjektumtól a leibnizi monász felé mozdult el”. A tanulmány ezt követően áttekinti a transzcendentalitás és a testiség feszültségének megjelenését a husserli fenomenológiában, majd a megoldást a kinesztétikus tudat fogalmának bevezetésében és kidolgozásában találja meg Husserl szövegeiben.

Takács Ádám írása („*Szubjektív tárgyiség*”. *Husserl fenomenológiai elmélete a test megalapozó szerepéről*), hasonló husserli problematikából kiindulva, gyakran ugyanazon szövegekre hivatkozva fogalmazza meg saját interpretációját, melyben talán még az előző tanulmányhoz képest is hangsúlyosabbá válik a testiség szerepe Husserlnél. A szerző kiindulópontja a transzcendentális tudat és a világ mint tapasztalati horizont elemzésének elválaszthatatlansága a húszas évek Husserlénél, amiből tanulmányában a transzcendentális szubjektum koncepciójában bekövetkezett változásokat vezeti le. Takács interpretációja az előző Husserl-elemzésnél jóval hangsúlyosabb szerepet szán a *test* és az *észlelés* husserli fenomenológiai elmélete rekonstrukciójának, és ebből kiindulva írja le, hogy hogyan válik Husserlnél a test az objektiváció forrásává általában: „A test fenomenalitása éppen arra a kérdésre adja meg a választ Husserlnél, hogy a *szubjektivitás eredendő önmegjelenése* miként képes motiválni az *eredeti tárgyi megjelenések* feltételeit”. Így válik interpretációjában a testtapasztalat elemzése többek között Husserl „metafizikájának” is az alapjává, a mai fenomenológia számára is érvényes nézőponttá.

Jóó Mária *A feminista elmélet és a (női) test* című tanulmányának deklarált célja a feminista elmélet filozófiai részének és az általában vett magyar filozófiai diskurzusnak a közvetítése egymás felé. A tanulmány rendszerezve ismerteti a feminista elmélet utóbbi néhány évtizedében fölmerült fontosabb nézeteket a (női) test témájáról. A szerző kategorizálása szerint a témában a posztstrukturalista felfogás (Michel Foucault hatása, Judith Butler, Susan Bordo) mellett az egyik leghangsúlyosabb jelenség a fenomenológiai módszer alkalmazása (Merleau-Ponty hatása, Iris Marion Young) a feminista elméletnek abban a külön

területként számon tartott részében, amelyet általában *corporeal feminism*ként emlegetnek. A tanulmány történeti áttekintése visszanyúlik egészen a *sex* és a *gender* megkülönböztetésének kezdeteiig (Simone de Beauvoir), majd az utóbbinak régebben nagy népszerűsége miatt tett diszkurzív konstrukcióként való interpretációjáig. Ebből az állapotból kiindulva lesz fontos, hogy később a „feminológiai alapokhoz jobban kötődő feminista kutatások az eltűnő testet igyekeztek visszahelyezni a centrumba”. A megélt test tapasztalatára építő feminista kutatások így találják meg azután a különböző emberi mozgásokat, az öltözködés testi tapasztalatát és az érzékelés és az érkezés szféráját mint (akkor) új kutatási területeket, egészen a *gender*-fogalom revíziójáig, néha megkérdőjelezéséig.

Moldvay Tamás *A spinozai test mint probléma* című tanulmánya annak a problémának a vizsgálatába vág bele, főként az *Etika* szövegére támaszkodva, „hogyan vajon mi is a test Spinozánál”. A problémafölvetés azonban korántsem tisztán filozófiatörténeti, hanem szorosan kapcsolódik a közelmúlt főként francia, a testtel kapcsolatban rendre Spinozára hivatkozó diskurzusához, jelesen Deleuze nézeteihez. A tanulmány első része a testet mint egyedi dolgot és mint individuumot vizsgálja, a kettő azonosíthatóságának problematikus voltára és a kauzalitással kapcsolatos összefüggésekre öszpontosít. Erre építve a szerző áttér a test és a mozgás kapcsolatának taglalására, s *a testet mint mozgást és mint a mozgások arányát* mutatja be Spinoza elméletében. Végigtekinti, hogy a mozgások *ratio*ja révén milyen típusú testek állhatnak elő, az egyneműnek tekinthető kőtől kezdve a világosan több részre tagolható testeken (halraj) és a mechanikus szerkezeteken keresztül (óra) az emberi testig, majd az állam testéig és a természet testként fölfogott egészéig. Az ilyen módon fölfogott testekkel kapcsolatos legfontosabb probléma a különböző *kimérák*, azaz testnek látszó, de sajátos mozgástörvényük szerint mozogva tartósan fennmaradni nem képes entitások lehetősége. A tanulmány gondolatmenete itt, a konklúzióban kapcsolódik újra az elején fölvezetett kortárs diskurzushoz, amennyiben a spinozai értelemben vett *kimérák*at a biotechnológia révén előálló entitások értelmezésében, leírásában gondolja hasznosíthatóknak. Másik példája, a kimérává váló államtest, szorosan kapcsolódik a következő tanulmány gondolatmenetéhez.

Balogh László Levente *A Lexiatán anatómiája* című írása a kora újkori politikafilozófia testmetaforáinak klasszikus példáját választotta elemzése tárgyául. Jóllehet a szöveg tisztán történeti elemzésként is megáll, a szerző elsősorban mégsem történeti érdeklődéssel közelít a szöveghez. Tanulmánya a hatalmi reprezentáció metaforáinak értelmezéséből indul ki, melynek csupán egy – habár alapvető, fontos és a továbbiakra nézve nagy hatású – történeti példája Hobbes fő műve. A szerző a hatástörténetből főként a huszadik századi német értelmezéstörténetet emeli ki, ezzel is jelezvén, hogy a kora újkor nem annyira önmagáért, mint inkább máig érvényesülő hatásai okán keltette föl érdeklődését. *A Lexiatán* államtest-metaforájának részletes, értő elemzését követően a szerző

végkövetkeztetésként a *rend* fogalmának átalakulása kapcsán annak az állapotnak a hobbesi megragadását regisztrálja, „amikor a rend már nem örök és eleve adott, hanem emberi eredetéből adódóan törekeny és állandóan újratemtendő lesz”. A tanulmány zárómondata szerint „kénytelenek vagyunk tudomásul venni, hogy a rend kapcsán nem az állapot örök, hanem maga a feladat”. Ez annak a jelenbeli állapotnak a leírása, amelynek megvilágításáért a szerző érdemesnek látta visszanyúlni az európai modernitás egyik forrásszövegéhez.

A tanulmányrovat záró darabja Gulyás Péter *A szimpátia és a belénkszületettség testi értelmezése Descartes-nál* című írása. A szerző Descartes-nak a filozófiatörténetesek többségének szemében másodrendűnek számító fiziológiai írásait állítja elemzése középpontjába. Az írás fő célja annak a bemutatása, hogy Descartes hogyan magyarázza az anya agyában lévő képeknek a magzat testére való rányomódását, és ennek fiziológiai leírása hogyan nyer metafizikai jelentőséget a karteziánus elméletben. Előbb azt vizsgálja a *Traité de l'homme* szövegének alapján, „hogy miként működik a testi érzékelés és az emlékezet, és hogy hogyan bírják az érzékelés és az emlékezet képei mozgásra a testet”, majd a *Description du corps humain* nyomán a karteziánus embriológiát mutatja be, ezt követően az anya agyának és a magzat testének kapcsolatára tér ki részletesen, amelynek a karteziánus elmélet mechanisztikus magyarázatot ad. A szerző írása utolsó részében a *szimpátiának* az anya és a magzat kapcsolatában betöltött szerepét vizsgálja, megmutatva, hogy a descartes-i fiziológia kérdései új szempontokat kínálnak az olyan klasszikus problémák megoldásához is, mint a test–elme viszony, az állati lélek létezése vagy az emberi léleknek az egész testben való szétterjedtsége.

A *Fórum* rovat tanulmány értékű, *Analitikus filozófia és fenomenológia* című vitacikke Eszes Boldizsár és Tózsér János tollából a magyar analitikus filozófusok és fenomenológusok között immár fél évtizede zajló vita folytatása, reflexió Ullmann Tamás *Analitikus és kontinentális filozófia* című írására. A szerzőpáros igyekszik kimutatni: Ullmann gondolatmenetének minden pontján az analitikus filozófia szűk, egy-két irányzatra korlátozott fogalmát használja, így megállapításai az analitikus filozófia előföltevéseiről, a kontinentális filozófia ezzel szemben fölvázolt jellemzőiről, valamint a két irányzat majdan megvalósuló egyesülésének jellegéről és lehetőségéről hamisak, vagy legalábbis kétségesek. A vitacikk fontos részei olvashatók a lapszám első tanulmányának ellenpontjaként is, habár az írások szerzői nem egymással vitatkoznak, és csak most, nyomtatásban ismerhetik meg egymás írásait. Sutyák Tibor elsősorban a fenomenológus szemével hasonlítja össze az analitikus filozófia és a fenomenológia testről alkotott nézeteit, míg itt – gyakran ugyanazokra a szövegekre hivatkozva – az analitikusok fenomenológia-értelmezését olvashatjuk több, gyakran a testtel kapcsolatos vitatéma kapcsán.

A *Szemle* rovatban Borbély Gábor hiánypótló középkori filozófiatörténetéről olvashatunk ismertetést. A kötetnek a tiszta intelligenciákra utaló címe – *Civakodó angyalok* – messze esik a *testtel* kapcsolatos problematikától, a mű mégis szá-

mos olyan kérdést és elméletet tárgyal, főként nyelvfilozófiai és ismeretelméleti megközelítésben, amelyekhez hasonlók későbbből is ismerősek a test-problematika körében is, de amelyekről mint történeti előzményről a későbbi korok kutatói és a kortárs filozófia művelői gyakran megfeledkeznek. (Elég itt példaként a mentális objektumok természetéről folyó vélekedéseket megemlíteni.)

Áttekintve a lapszám szövegeit, feltűnő, hogy a szerzők a *test* problémájának különböző szempontú taglalása során rendre az oksággal összefüggő problémákba ütköztek. (Például a különféle mozgások magyarázata során, vagy az igen gyakran visszatérő *mentális okozás* kérdése kapcsán.) Nem csupán a véletlen műve, hogy a *Magyar Filozófiai Szemle* következő rendes számának témája éppen a *kauzalitás*.

*Mester Béla*

## A test–test probléma

Sok baj van a testtel. Az utóbbi évtizedek filozófiai reflexiójában ez a szó – test – legalább két igen markáns kutatási és elméletképzési tradícióban, az analitikus elmefilozófiában és a fenomenológiában a problémátér egyik legsúlyosabb eleme, ráadásul más és más szempontból. Egyik oldalon egy masszív rejtély megszerkesztését, a másik oldalon egy klasszikus kérdés korántsem megnyugtató elnyelését tapasztaljuk. A maga módján mindkettő tarthatatlan – mégis kitartó.

Kezdjük a rejtéllyel. Amiként az utóbbi időben többen is leszögezték (Kim 2005; Ellis 2000; Hanna–Thompson 2003) – csak hogy olyan szöveget szemlélzünk, amelyek egyébként nem sokban egyeznek –, az elmefilozófia ágas-bogas témaszövevénye, amely az intencionalitástól a propozicionális attitűdökön át a másik elme kérdéséig hatalmas adag, technikailag művesen kidolgozott filozófiai anyagot fog át, végeredményben két, a test és az elme viszonyát firtató problémába sűríthető: az egyik a fenomenális minőség ügye, a másik pedig a mentális okozása. A fenomenális minőség vagy más néven fenomenális tapasztalat (a tapasztalat fenomenológiai karaktere, szubjektív tapasztalat, szubjektív nézőpont, kvália stb.) azért problematikus, mert noha valamennyien rendelkezünk vele, mi több, sajátlagos hozzáférésünk van hozzá, és intuitív pszichológiai reflexiónk szerint kulcsszerepet játszik viselkedésünk motivációs hátterében, sehogy sem tudjuk megmagyarázni, hogy miképpen áll elő testünk neurológiai-fizikai felépítéséből következően. Vagyis noha ténynek gondoljuk, hogy a kávéillat bizonyos gázmolekulák és bizonyos szaglószervi receptorok, továbbá ingerület-átvivő kemikáliák és más anatómiai egységek oksági kapcsolatában áll elő, és ugyancsak ténynek gondoljuk, hogy egy adott helyzetben a kávéillat a kellemes várakozás fenomenális karakterét mutatja fel számunkra, a két tény között nem vagyunk képesek sem oksági, sem funkcionális kapcsolatot teremteni. Hogyan lesz a kémiai folyamatból élvezet, várakozás, undor vagy bármi más szubjektív tapasztalat? Hogyan éri el az agyunk és az idegrendszerünk, hogy affektív minőségeket éljünk át? Hogyan lehetne olyan magyarázati stratégiát találni, amely gördülékenyen, zökkenés- és ugrásmentesen térne át a fizikai-strukturális as-

pektusról (molekulák, afferens idegi huzalozottság, neurotranszmitterek stb.) a megélt-fenomenális aspektusra (illat, vágódás, kellemes íz-előérzet stb.)?

Amennyiben fenomenális tapasztalataink logikailag ráépülnek agyi-idegi állapotainkra, akkor egyrészt egy meghatározott agyi-idegi állapot fennállása szükséges és elégséges feltétele lenne a hozzá tartozó meghatározott mentális állapot fennállásának, másrészt a meghatározott agyi-idegi állapot ismerete elégséges feltétele lenne a hozzá tartozó meghatározott mentális állapot ismeretének. De a dolog úgy áll, hogy bármikor elgondolható olyan helyzet, amelyben fennáll a meghatározott agyi-idegi állapot, ám a hozzá tartozó fenomenális állapot nem áll elő (Putnam 1975; Chalmers 1996). És bármikor elgondolható olyan helyzet, amelyben beható ismereteink vannak egy meghatározott agyi-idegi állapot szerkezetéről, még sincs a leghalványabb fogalmunk sem az – állítólag – hozzá tartozó fenomenális karakterről (Nagel 1974/2004; Jackson 1982). Ebből következően sem ontológiai, sem episztemológiai szinten nem áll, hogy az agyi-idegi állapotok minden további nélkül maguk után vonnának meghatározott fenomenológiai állapotokat. Ha pedig ez nem áll, akkor az sem jelenthető ki, hogy a fenomenológiai állapotok levezethetők volnának az agyi-idegi állapotokból, fogalmazódjon meg a levezetés akár az azonosság, akár a redukció, akár az okság, akár a realizáció terminusaiban.

A nyolcvanas évek elején volt, aki úgy gondolta, hogy a privát minőségi tapasztalat kérdése mint akadály elhárult a specifikusabb elmefilozófiai problémák megoldása elől (Churchland 1981). Időközben azonban a kérdés sokkal makacsabbnak bizonyult. Ha konszenzus nincs is, erőteljes megfontolások szólnak amellett, hogy a kvália problémája kikényszeríti a következő belátást: a fenomenális mentális állapotok és események minőségileg eltérőek a fizikai állapotoktól és eseményektől, nem vezethetők vissza azokra, és nem magyarázhatóak a segítségükkel.

A mentális okozás témája csak tovább bonyolítja a helyzetet. Míg ugyanis a tapasztalatok fenomenológiai karakterének problémája elzárta ugyan az utat az egyszerű típusazonossági hipotézis előtt (vagyis az előtt, hogy meghatározott fenomenológiai állapotaink nem egyebek, mint meghatározott fizikai állapotok – a fájdalom nem más, mint a C-rostok tüzelése, szól a fájdalmasan agyonhasznált példamondat), a példányazonosság lehetősége még nyitva maradt (egy kvalitatív állapotot mindig egy fizikai állapot realizál vagy instanciál). A példányazonosság tulajdonságdualizmust implikál: ha egyszer a fenomenális és a fizikai állapot szubsztanciálisan azonos, akkor valamiben mégis különbözniük kell, mégpedig nem egyebben, mint a tulajdonságok dolgában. A létezőknek tehát vannak fizikai és vannak kvalitatív öntapasztalásként megjelenő tulajdonságai. Utóbbiak nyilván mentális tulajdonságok. Abban van ugyan vita, hogy a mentális tulajdonságok mindegyike egyben fenomenális tulajdonság-e, azaz hogy megjelenik-e fenomenálisan hordozójának, illetve ha megjelenik, feltétlenül kvalitatív jellegű megjelenést kap-e, és noha ez a vita a legkevésbé sem érdektelen, most, az

egyszerűség kedvéért, az általában vett mentális tulajdonságok kapcsán fejtem ki a mentális okozás problémáját – mintha a nevezett vitát fű alatt eldöntöttem volna. Később visszatérek a szűkebb fogalmiságra; egyébként a rekapitulálandó érvet ez a kérdés nem érinti.

A tulajdonságdualizmus elfogadása esetén tehát azzal kell szembenézni, hogy vajon mennyire vagyunk képesek valódi, azaz effektív pozíciót biztosítani a mentális tulajdonságoknak praktikus életünkben. Hajlamosak vagyunk azt gondolni, hogy mentális aktusaink okságilag hatékonyak, tehát például, ha kávéét kívánok inni, aztán pedig iszom kávéét, akkor megkívánásom okságilag motiválta cselekvésemet, vagyis egy mentális történés kiváltott egy fizikai történést. Csak-hogy, mint Jaegwon Kim ismétlődően kifejtette (Kim 1993; 2005; 1993/2008), ez az intuitív elgondolás olyan kiegészítő feltevéseket működtet, amelyek együttesen már nehezen tarthatók. A tulajdonságdualizmus éppen azért az, ami, mert nem kívánja meg a szubsztanciák dualista felosztását, vagyis egy fizikai és egy mentális szubsztancia feltételezését. Szubsztancia-monista álláspont ez, még hozzá egy fizikalista ontológia talaján. Elvileg persze beszélhetnénk tulajdonságdualizmusról egy mentalista monizmus alapján is, tehát egy pánpszichista álláspontról, ám ez nagyon nehézkes metafizikai kereteket igényelne. Van ilyen nézet (lásd Seager 1995/2004), és ambíciói között szerepel is a mentális okozás problémájának megoldása, de megfogalmazásához a matematikai információelmélet és a kvantummechanika nagyon specifikus filozófiai interpretációjára van szükség, ami beláthatatlan messzeségekbe tolja ki a teoretikus elmélkedést. A tulajdonságdualizmusról tehát mint fizikalista álláspontról beszélek, amely szerint a világban csak olyan dolgok (entitások és események) léteznek, amelyeket a fizika (vagy egy idealizáltan *teljes* fizika) tudománya tárgyal. A mentális tulajdonságok (események, aktusok) is fizikailag instanciáltak. Mármost bármi történik, annak – tekintve a fizikalizmus ontológiáját – oksági eredete a fizikai dolgok fizikai törvények szerinti viselkedésében áll, és – mivel a fizikai dolgok fizikai törvények szerinti viselkedése minden fizikai változásra elégséges alapot biztosít (a fizikai világ oksági zártságának elve szerint) – nem áll semmi másban. Az, hogy „semmi másban”, egyben azt is jelenti, hogy nem fogadjuk el a változások oksági túldetermináltságának lehetőségét, azt tehát, hogy egy egyedi eseményt több egymástól független elégséges ok eredményez. A probléma ekkor az, hogy én megkívánom a kávéét, megiszom a kávéét, ám az, hogy megkívántam, fizikailag instanciált esemény, vagyis van egy strukturális (agyi-idegi) hordozója, amely önmagában okságilag hatékony kell legyen a kávéivás fizikai eseményét illetően – különben valamilyen fizikán túli instanciát kellene feltételeznem, amit viszont kizár az elfogadott elméleti alapállás. Ha tehát fizikai strukturák okozzák a fizikai strukturák átrendeződését, és egyúttal kizárjuk azt, hogy ugyanilyen hatékonysággal valamilyen más típusú esemény vagy jelenség *is* oksági szerepet játsszon az ügyben (túldetermináltság), akkor egyszerűen nem tudjuk megmondani, hogy a megkívánásnak mint mentális epizódnak mi a helye a kávéivás történetében.



Tényleg nagy a baj. Az, hogy kávét kívánok, csak akkor vezethet kávéiváshoz, ha a szóban forgó mentális állapotom *fizikailag* okozza a kielégítő fizikai állapotot, fizikailag viszont csak akkor képes ilyesmit okozni, ha maga is fizikai természetű. Igen ám, de éppen azért fogadtuk el (ideiglenesen) a tulajdonság-dualizmus álláspontját, mert valamiképpen azt igyekeztünk biztosítani, hogy a mentális állapotokat *ne* kelljen fizikai állapotokkal azonosnak (legfeljebb fizikailag realizálnak vagy instanciálnak) gondolni. Valójában felértékeljük a mentális állapotokat azáltal, hogy legalább részleges autonómiát biztosítottunk nekik a fizikai állapotokhoz képest (ezért is szokás ezt az elgondolást „nem reduktív fizikalizmus”-nak becézni, mert hát nem akar reduktív lenni), de, mint kiderült, mindezt annak árán tettük, hogy megfosztottuk őket oksági hatékonyságuktól. Ez az úgynevezett „kizárási érv” összesen két lehetőséget hagy nyitva a mentális okozás ügyében: vagy azt fogadjuk el, hogy a mentális állapotok epifenomenálisak, vagy azt, hogy mégis csak azonosak fizikai állapotainkkal.

Az első esettel az a gond, hogy a mentális állapotok epifenomenalitása nyitva hagyja ugyan azt a lehetőséget, hogy ezek az állapotok (események, tartalmak) nem vezethetők vissza fizikai állapotokra (eseményekre, struktúrákra), csak éppen édes mindegy, hogy ki, hová és mihez vezetí vagy nem vezetí őket vissza, ha egyszer *annyira* szánni valóan hatástalanok az életünkre. Afféle paszományai csak a fizikai létezésnek; nem valók arra, hogy összekössenek bármit is. Az epifenomenalizmus azonban egyrészt alapvető intuíciónknak látszik ellentmondani – ha kávét kívánok, iszom kávét, és azért iszom, mert megkívántam (ha minden flottul megy) –, másrészt nyitva hagy egy különös lehetőséget, nevezetesen azt, hogy lehetségesek lennének hozzánk megtévesztően hasonló viselkedésű lények, akik funkcionálisan mindenben olyanok volnának, mint mi, csak éppen nem volnának belső mentális állapotaik. Hiszen ha egyszer a mentális tartomány nem avatkozik bele a világ működésébe, akkor elvileg nélküle is minden ugyanúgy menne a maga útján, és a különbség észrevehetetlen volna. Visszatérünk még ahhoz a kérdéshez, hogy ilyesmi egyáltalán elgondolható-e.

A második eset – hogy mentális és fizikai egyszerűen ugyanaz – viszont a fentebb emlegetett megfontolások miatt kevésbé elfogadható. Ha minden mentális állapot azonos egy megfelelő fizikai állapottal, és a kvalitatív fenomenális tapasztalat mentális állapot, akkor az is azonos egy megfelelő fizikai állapottal. De ha azonos vele, akkor, mint szó volt róla, hogyan lehet különféle fizikai struktúrák révén instanciálható, és hogyan lehetséges az, hogy fizikailag mindent tudunk róla, fenomenális sajátjaihoz mégsem kerülünk közelebb? A tapasztalat fenomenális karakterének és a mentális okozásnak a kérdése együttesen valódi filozófiai rejtélyt generál, amelyben az egyik szál felgöngyöltése, úgy tűnik, a másik szálát csak jobban összekuszálja.

Szembeszökő, hogy ebben az egész elmefilozófiai kérdéskörben a *test* terminusa valójában az agyi és idegrendszeri struktúrákra korlátozódik. Ez kétségtelül erősen szűkítő fogalomhasználat: a test és a lélek klasszikus problémáját az

agy és az elme kapcsolatának témája már eredendően redukálja. Ennek a redukciónak a magyarázata egyfelől az elmefilozófiai tradíció természettudományos elkötelezettségében keresendő – jelen esetben ez a természettudomány első sorban a neurológiai kutatásokat jelenti. Másfelől ezt a fogalomkezelést erősíti a Kripke és Putnam alapvető szövegeiben körvonalazódó és igen nagy tekintélyre szert tett externalista szemantika (Kripke 1972/2007; Putnam 1975/2010), vagyis az a gondolat, hogy természeti fajtaikat jelölő terminusaink jelentését végső soron nem az elme tartalmai, hanem a dolgok fizikai mikrostrukturái rögzítik (na meg az eredendő megnevező aktusok és a mindenkori szakértői konszenzus). A dolgok fenomenális aspektusaiért fizikai összetételük felelős, a víz átlátszósága, viszkozitása, forrás- és fagyáspontja és egyéb ismert sajátosságai a hidrogén- és oxigénatomok kémiai tulajdonságaira vezethetők vissza. A mi világunkban az ezeket a fenomenális sajátosságokat felmutató anyagot a  $H_2O$  mint kémiai struktúra instanciálja, így a „víz” kifejezésünk jelentése ebben a szerkezetben rögzül. Ebből a nézőpontból akár úgy is megközelíthető a kérdés, hogy az ilyen és ilyen fenomenális tulajdonságokat felmutató (kvalitatív tapasztalat, intencionalitás stb.) „elme” a mi világunkban szintén visszavezethető kell legyen valamiféle fizikai mikrostrukturára. A bökkenő az, hogy az elme fenomenális tulajdonságai elsősorban magának az elmének tárulnak fel, tehát a külső megfigyelő nézőpontját (a víz ilyennek látszik, de ha jobban megnézem, kiderül, hogy pontosan miért ilyen) a belső, önátélő szubjektivitás nézőpontjával kell kiegészíteni, sőt valójában ez utóbbinak a tapasztalata képezi az elmefilozófiai rejtély volta-képpeni tartalmát.

Éppen ez a körülmény motiválja a „magyarázati szakadék” vagy „magyarázati rés” meglétét tematizáló megfontolásokat (Nagel 1974/2004; McGinn 1989/1999; Levine 1983). A természettudomány, mondja Nagel, alapvető módszertani beállítottságának megfelelően, eltekint a szubjektív fenomenális tapasztalattól, hogy a jelenségek és események olyan szerkezeti rekonstrukciójához jusson, amely egyrészt minden partikuláris nézőponttól független, másrészt képes arra, hogy oksági magyarázatot adjon bármely partikuláris nézőpont fenomenális tartalmára. A szubjektív tapasztalat ügyében azonban ez a módszertani gyakorlat a tárgy eltüntetését jelenti, hiszen amennyiben eltekintünk a fenomenális aspektustól, miközben vizsgálódásunk tárgya éppen a fenomenális aspektus, vizsgálódásunk egyszeriben azt sem tudja, mit is vizsgáljon. Nem vitás, teszi hozzá McGinn, hogy mentális életünk valamiképpen agyi-idegi életünk szerkezetén múlik, ám a lényeg pontosan az, hogy ezt a „valamiképpen”-t nem vagyunk képesek határozott „így és így”-re váltani egy koherens beszédmódon belül, hiszen a neurológiai megközelítés térbeli lokalitásokat, a fenomenológiai reflexió pedig kvalitatív (azaz: nem térbeli) tartalmakat tárgyal. Ha létezik is a nagy, áthidaló elmélet, az emberi gondolkodás számára – nyelvi és módszertani korlátok miatt – hozzáférhetetlen marad. A mikrostrukturákat azonosító idegtudományos magyarázatot és a belső élmények megragadását nem lehetséges szinkronizálni.

Csakhogy a test nem okvetlenül azonosítandó az agyi-idegi mechanizmusokkal, még az elme vagy a tudatos tapasztalat filozófiai problémájának kontextusában sem. A Husserl nyomán kibontakozó fenomenológiai vizsgálódások a test olyan elgondolása előtt nyitották meg az utat, amely egyrészt soha nem korlátozta az emberi test fizikai fogalmát az agyra vagy az idegrendszerre, hanem egy egészes korporeális fenomént tematizál, csontszerkezettel, izmokkal, morfológiai kényszerekkel; másrészt pedig ettől a fizikai testtől elválasztja a megélt, önmagának megjelenő test tapasztalatát, amelynek konstitutív aktivitása a világban való tájékozódás és boldogulás kulcsa. Mint ismeretes, Husserl a *Körper* és a *Leib* fogalmaival jelölte a szóban forgó distinkciót (Husserl 1989). A test egyfelől megjelenik a világban mint bármely más dolog, aspektusok és szintézisek sorozatában áll össze egységes fenoménné, másfelől viszont önmaga ennek a fenomenalizációnak az alanya, mozgásával, cselekvési lehetőségeinek kamatoztatásával folytonosan újraképi önnön világát, és belső, proprioceptív és kinesztetikus tapasztalata van önmagáról. Ezt az alapvető belátást vette át és radikalizálta Merleau-Ponty, amikor az észlelésről – értsd: a világ megjelenéséről – írott nagyszabású munkájában (Merleau-Ponty 1945) a saját alkata és a környezete kölesönös egymásra-vonatkoztatottságában szituált testet tette meg a világot feltárva megértő emberi egzisztencia realizációjának. A test ilyképpen előáll kettős tapasztalata – fizikai test egyfelől, megélt testtapasztalat másfelől – nyilvánvalóan nem esik egybe. *Tudásom* arról, hogy a koffein micsoda élénkítő folyamatokat gerjeszt az idegrendszeremben, és *érzésem* arról, ahogy kicsit elfordítom a fejem, hogy a kávéillatot még intenzívebben érezhessem, a harmadik személyű megfigyelés és az első személyű megélés elválása egymástól. A tapasztalatoknak ez a szétartása a test-elme problémán túl (vagy ahelyett) egy test-test problémát generál: miképpen kapcsolódik egymással össze a fizikai és a fenomenális test?

Szubsztancia-dualizmusról nyilván ezúttal szó sem lehet. Husserl és Merleau-Ponty közösen osztott és mindkettejük által alaposan elemzett észrevétele, amely szerint a saját test olyan kitüntetett tárgyi létező, amely nem távolítható el tőlem, nem sokszorozhatom meg körüljárással észlelt aspektusait, mind a reális, fizikai testre, mind a tapasztalásban átélt *Leibra* érvényes. A testem mint a világ része és a testem mint a világ énpólusa ugyanaz a test, ám megnyilatkozásaik különbözőek. De vajon a fizikai testem realizálja-e mindazt, ami testi önátélésben előkerül, vagy fordítva, megjelenő testem egyik, speciálisan modifikált fenomenális aspektusa volna strukturális felépítemsem? Esetleg két egymásra visszavezethetetlen tulajdonságszféra megnyilvánulása egyik és másik? És ráadásul: hogyan kapcsolódik ehhez a kétarcú jelenséghez, a testemhez az, amit elmének vagy tudatnak neveznek?

Ezek a kérdések úgyszólván maguktól adódnak. Annak az okai, hogy a fenomenológiai tradíció klasszikusainál mégsem találunk pontos és határozott választ rájuk, magában a beállítottságban és a módszerben le lehetők fel.

Husserl számára a test fenomenéje egyfajta „fenomenológiai anomália” módjára tematizálódott, hiszen nem illeszkedett sem az immanens tudatfolyam és a transzcendens tárgyiságok, sem a reális dolgok és az ideális lények megkülönböztetéséhez (Carman 1999). Közbülső és közvetítő elemként az anyagi világ és a szubjektív szféra között a test a szubjektumhoz tartozó lokalizációs mezőként értelmeződik. Az érzéki modalitások fenomenológiai aszimmetriájának vizsgálata a taktilis modalitás sajátosan önreflexív karakterének feltárásához vezetett – a látó szem nem látja magát, miközben lát, a halló fül nem hallja magát, de a tapintó kéz önmagát is érzi a felület megérintésekor –, és ez az önreflexivitás a testi mozgásos konstitutivitás analízisét eredményezte, ám mindez alárendelődött a transzcendentális ego spontaneitása faktumának. Husserl álláspontja a hagyományos test-lélek kérdésben a *Leib* kapcsán sem mond búcsút annak az esszenciadualizmusnak vagy kétaspektus-elméletnek, amely szerint ez ügyben egy azonos hordozó (az én, a szubjektum) két lényegeszerű mozzanatáról van szó. Egyik oldalon van a *természet*, azaz a tér-időbeli realitások oksági viszonyok szerint működő sokasága, a másik oldalon pedig a *tudat*, a maga nem térbeli, de időbeli tartalmaival, amelyek nem az okság, hanem a motívum és az intencionalitás szerint szerveződnek. David Woodruff Smith a test és az elme viszonyának husserli megközelítését Davidson anomáliás monizmusához hasonlította (Smith 1995), ami vitatható megközelítés ugyan (Husserl soha nem fogadta volna el a nomologikusan kauzális fizikai szféra episztemológiai elsőbbségét), de kifejező: mindenképpen egy ontológiai monista, ám módszertanilag nem reduktív elmélet az, ami egyáltalán szóba jöhet egy efféle rekonstrukcióban.

Merleau-Ponty az egész elmefilozófiai problémakört egy teoretikus hagyomány származtatott termékének tartotta. Ahogy *A viselkedés szerkezete* zárófejezetében írja (Merleau-Ponty 1942), a természetes tapasztalat nem osztja fel az emberi egységet lélekre és testre, a test nem érzéketlen gép, és a lélek sem cselekvésünk és gondolkodásunk kísértetszerű parancsnoka. A fenomenológiai leírásnak oda kell utat találnia, ahol a közvetlen és prereflexív testi tapasztalat részt vesz a világ születőfélben lévő logoszának kialakításában, vagyis a testileg megvalósuló értelemképzéshez. Innen nézve pedig (már ha sikeresen reflektálunk a prereflexívra – ez a kérdés Merleau-Ponty egész filozófiai életművének fókuszában áll) a test és az elme különbsége, a kettejük között lévő oksági hatékonyság kérdése mesterséges fogalmi oppozíciók létesítéséből ered. Nem azt kell kérdezni, hogy egy tisztán mentális aktus (fel akarom emelni a kezem) hogyan képes kiváltani a fizikai mozgást (felemelem a kezem), hiszen a testemről szerzett eredendő tapasztalatomban cleve benne van az, hogy képes vagyok felemelni a kezem. A testi önmegélés alaptapasztalatai más státusúak, mint az úgynevezett népi pszichológia úgynevezett elméletei, amelyek, ha hamisnak bizonyulnak, ugyanúgy elvethetőek, mint az égés flogisztonelemlélete vagy az élet vitalista koncepciója. Nem is elméletek ezek, hanem a világban élő emberi lény

inkarnált lehetőségei. Minden teória már csak ezek *után* és ezek *eredményeképpen* képződhet meg.

Csakhogy az, hogy egy probléma származékos, kontingens, esendő gondolati tradíciók fogalomképzése nyomán motivált, még nem jelenti okvetlenül azt, hogy egyben irreleváns is. Elfogadva azt a pozíciót, amit a testnek a röviden szemlézett fenomenológiai belátások tulajdonítanak, hogy tehát nem egyenlő az aggyal és az idegrendszerrel, hogy maga is két elkülöníthető aspektusra – egy fizikai-strukturális és egy megélt-fenomenális aspektusra – bomlik, a csavaros rejtély továbbra is fennáll: miképpen áll elő ez a bizonyos fenomenális aspektus, mi köze van a fizikaihoz, és miképpen hatnak egymásra életünkben?

Körülbelül két évtizede került forgalomba – vagy legalábbis vált gyakran használatossá – egy kifejezés, amely egyben Francisco Varela, Evan Thompson és Eleanor Rosch programadó könyvének (Varela–Thompson–Rosch 1991) is a címe: *embodied mind*, azaz „megtestesült elme” vagy „testet öltött elme”. A szóösszetétel hamar meglehetősen nagy karriert futott be, és olyan, egymással csak laza kapcsolatban álló kutatások címkéjévé vált, mint amilyen George Lakoff és Mark Johnson munkái a kognitív szemantika területén, illetve ennek kiterjesztése a filozófiai elmélkedés területére (Lakoff–Johnson 1999; Johnson 2007); Andy Clark elgondolásai a technikai környezet és az intelligencia viszonyáról (Clark 1997; Clark–Chalmers 1998); Rodney Brooks hasonló témájú vizsgálódásai a mesterséges intelligencia ügyében (Brooks 1991; 1999); Hubert Dreyfus és Shaun Gallagher filozófiai reflexiói a fenomenológia és a modern viselkedéstudomány eredményeinek összekapcsolhatóságáról (Dreyfus 1992; 2002; Gallagher 1995; 2005; 2007). Természetesen megszülettek az ilyenkor szokásos gyűjteményes kötetek (például Bermúdez–Marcel–Eilan 1995) és összefoglalók (Gibbs 2005; Anderson 2003). Magyarul is napvilágot látott néhány olyan szöveg, amely áttekinti az irányzat főbb tematikus egységeit, eszmetörténeti kötődéseit és perspektíváit (Kampis 2001; 2002; Szigeti 2002). Ám, ahogy az gyakorta előfordul, a szövegbőség nem feltétlenül segít a teoretikus rend fenntartásban. Érdemes néhány fogalmi észrevétellel előhozakodni, azután pedig tisztázni, hogy miért került itt most elő ez az irányzat, mi köze az elmondottakhoz, és mit lehet kezdeni vele.

Először is: a „testi elme” kifejezést fogom használni az *embodied mind* megfelelőjeként, és a „testiség”-et az *embodiment* magyarítására. Azért nem a szó szerinti „megtestesült elme” és „megtestesülés” vagy „testet öltött elme” alakok mellett döntöttem, mert egyfelől tanácsosnak gondolom kizárni ezeknek a formuláknak a teológiai konnotációját (amit egyébként az angol verzió, sőt az annak alapjául szolgáló francia *incarnation* sem zár ki teljesen, de ez legyen az ő gondjuk); másfelől a „megtestesülés” azt sugallja, hogy az elme mintegy *utóbb* költözött volna a testbe, tehát, hogy volna egy önmagában álló, testetlen elme, amely aztán valamiképpen bekerül egy testbe. Ezzel szemben a testi elme kon-

cepciója éppen azt tartalmazza, hogy az elme eredendően és lényegileg testi természetű, működése testileg kondicionált, és nem lehet róla másmilyen, mint testi tapasztalatunk.

A testi elme gondolata számos elméleti törekvésre támaszkodik. Hivatkozásai között gyakran megtalálhatóak Husserl és Merleau-Ponty testfönetenológiai belátásai, valamint Heidegger tanítása az önmagát megértő emberi egzisztencia világ-involváltságáról – a fönetenológiai hagyomány mint filozófiai erőforrás jelenik meg, kisebb részben konkrét megoldások átvételével, nagyobb részben a beállítottság kiaknázására intő programként. Explicit célkitűzése, hogy megoldást találjon az elme-filozófiai rejtélyre, méghozzá általában az eredeti elme-filozófiai problematika fogalmi átrendezésével, kimutatván a hagyományos megközelítés homogenizáló tendenciáinak merevségét (Clark–Chalmers 1998; Gelder 2005). Ugyanakkor a tisztán filozófiai elmélkedés mellett az irányzat empirikusan elkötelezett: rengeteget merít a kognitív viselkedéstudomány kísérleti beszámolóiból, pszichopatológiai esetleírásokból, és erőteljesen támaszkodik a kortárs neurológiai kutatások eredményeire. Látható, hogy a testi elme szintetikus koncepció: jóindulatú megfogalmazásban interdiszciplináris kötődésű, rosszhiszeműbben megközelítve eklektikus teoretikus talapzaton áll.

Ez a szintetikus (vagy szinkretikus) elméleti alapállás már a középponti fogalom, a *test* használati módjaiban is megnyilatkozik. A „test” fogalma ugyanis egyaránt kiterjed az agyi-idegi huzalozottság neurofiziológiájára, a világban tájékozódó és cselekvő, anatómiai lehetőségei által (is) meghatározott alkat viselkedésére, illetve az önmagát affektív módon megtapasztaló emberi test önfönetenalizációjára. Ez a háromrétegű testfogalom némely szerzőnél kissé zavaros egység-koncepcióként működik (Lakoff–Johnson 1999), másoknál azonban önálló problémafelvetés és magas szinten reflektált kontribúciók forrása (Hanna–Thompson 2003).

És ha már a fogalomhasználat ügyénél járunk: ahhoz, hogy higgadt mérleget készítsünk a testi elme gondolatában rejlő filozófiai lehetőségekről, el kell tekintenünk az elmélet kifejtésében hangot kapó mozgalmár, túlfűtött, időnként profétikus retorikától. Ez magától értetődőnek tűnik, de aki belc olvasta magát az irányzat alapszövegeibe, azt hiszem, egyet fog érteni abban, hogy a megszövegezés módja két ellentétes veszélyt is megjelenít. Félő, hogy az olvasó vagy túlságosan komolyan vesz bizonyos hiperbolikus elméleti kirohanásokat (hogy minek az ideje járt le, mivel kell holnaptól felhagyni, és hogyan változik meg rövidesen úgyszólván minden a reflexióban), vagy ezek miatt éppenséggel semmit sem fog komolyan venni a mondottakból. A józan befogadói disztíngváltság e helyt feltétlenül gyümölcsöző.

A testi elme gondolatának bemutatásakor szót kell ejteni (1) a reprezentáció-kritikáról; (2) a szituatív kogníció teorémájáról; (3) a mozgásos konstitúció témájáról és (4) az affektivitás kezeléséről.

1. *Reprezentáció-kritika.* Az észlelés és a cselekvés viszonyának hagyományos rekonstrukciója (amit a klasszikus kognitívizmus is átvett) a mentális képek vagy tartalmak központi, koordináló szerepére épített. Az elképzelés szerint a külvilág tárgyairól és eseményeiről az érzékelés révén mentális képzetek (ideák) formálódnak az elmében, a gondolkodás ezeket a képzeteket manipulálja, majd utasítást ad a legmegfelelőbbnek ítélt cselekvés végrehajtására. Ezt az egy-az-egyben leképezést a kognitív tudományban egy olyan szimbólum-feldolgozás koncepciója váltotta fel, amelyben a mentális (előhuzalozott vagy integrált) szimbólumoknak nem kell morfológiailag a külvilág tárgyaira hasonlítaniuk, ugyanakkor a meghatározott szimbólumok továbbra is meghatározott dolgok, tények helyén állnak. A gondolkodás a szimbólumok szabályszerű transzformációját jelenti, majd a kiválasztott megoldás visszakódolódik efferens cselekvési utasítások formájába. A folyamat tehát a következő fázisokból áll: érzékelés, modellalkotás, tervezés, majd kivitelezés. Csakhogy ez az egész processzus, mondják (Brooks 1991; Gelder 1995; Clark–Chalmers 1998), túlságosan nehézkes; egyszerűen lekésik a valós idejű problémamegoldásról. A legtöbb habituális cselekvésünk esetében nincs is fenomenológiai fedezete a modellalkotás és tervezés fázisának, más szóval semmi nem támasztja alá a reprezentációs alapú mentális mechanizmust. Ahogyan Brooks fogalmaz híressé vált tanulmányában (Brooks 1991), célravezetőbb a világot használni önnön modellje gyanánt, vagyis helyesebb egy olyan rekonstrukció, amely szerint a sztereotip cselekvésekkor maga a test, a végtagok, az izmok és ízületek végzik el a feladatot, nincs szükség központi feldolgozásra. Ez nem jelent visszatérést a behaviorizmus programjához, nem arról van szó, hogy kiiktatnánk az elme „fekete dobozát” a viselkedésből, inkább arról, hogy a „fekete doboz” nem kizárólag az agyban, hanem a kézben, a lábban, a derékban, az egész testben lokalizálendő. A kezem megtanulja, hogyan fogja meg a kávésesésze fülét, és ezt könnyen meg is teszi anélkül, hogy akár a csészét, akár a kezemet reprezentálni vagy modellezni kellene az elmében – a test diszpozicionális viselkedési begyakorlásáról beszélünk tehát.

Másrészt a reprezentációval kapcsolatban egy elvi probléma is felbukkan. Ha egy gyakorlati helyzet észlelési alapú modellezése folyik, akkor ennek a folyamatnak valamiképpen ki kell szűrnie a környezetből a szituáció szempontjából releváns paramétereket, és el kell azokat választania az irrelevánsaktól. Ám azt, hogy mi releváns és mi nem, csak akkor értékelhetem megfelelően, ha már felismertem, hogy mi az, amit a helyzet megkíván tőlem – viszont a helyzet felismeréséhez már el kellett választanom a relevánsat attól, ami nem az. Mindez nyilván nem jelent igazi problémát a helyzet *utólagos* mentális értékelésekor, de a valós időben gyakran irgalmatlanul gyorsan lefolyó akciók és reakciók áramában bénító hatású praktikus paradoxonként áll elő. Ezt a paradoxont tünteti el a testiség gondolatának pragmatikus hermeneutikája: mindig már benne vagyok egy helyzetben, csinálom azt, amit csinálok, és cselekvésem is már elkezdődött a számomra. Ennek a beágyazottságnak köszönhetően a testem ismeri fel azt, ami

számít – annyira fókuszálok a szememmel, annyira emelem fel a fejem, annyira zárom az ujjaimat, amennyire az éppen adott szituáció megkívánja.

Persze ez nem jelentheti a reprezentáció *teljes* kiiktatását a problémamegoldás folyamatából. Például ha elakad a cselekvéssor, ha korrekcióra van szükség, ha reflexíve át kell gondolni, mi is volna a helyesebb alternatíva, akkor explicit reprezentációkra, köztük a saját testről alkotott reprezentációkra és azok feldolgozására van szükség. De ezek is csak akkor működhetnek sikeresen, ha előzőleg már behelyezkedtem a szituációba, és ez a behelyezkedés mindig testi.

2. *Szituáltság.* A reprezentáció kritikája közvetlenül a szituáltság belátásához vezet át. Ez a filozófiailag Heidegger-től és Merleau-Ponty-tól származó gondolat azt mondja ki, hogy minden egyes, még a legimmanensebbnek felfogott mentális folyamat is mindig valamilyen konkrét, feladatként is értelmezhető helyzetben bomlik ki. Ha felkapom a fejem valamire, ha szórakozottan dudorászom, ha logikai problémákkal kínlódom, ha kávéra vágyakozom, minden esetben egy praktikus, elemeit tekintve egymással örökösen keresztutalások hálózatában álló világ mint környezet orientálja a magatartásomat, és magatartásom e hálózat nélkül nem is értelmezhető. Márpedig ez a szituáltság elsődlegesen és eredendően testi: testem pozíciója határozza meg, hogy éppen mit jelent az, hogy *itt* vagyok, hogy valaminek a *közepette* vagyok itt. Az észlelés azért nem lehet pusztán a külvilág belső reprezentációja, immateriális és passzív recepció, mert az, hogy éppen milyen a testtartásom, merre fordulok, *hol áll a fejem*, eleve megmunkálja, preformálja az alakuló észlelési tartalmat. A világ perspektivikus és dinamikusan változó vázlatok formájában adódik, és a mindenkori perspektíva nem más, mint saját testem látószöge. Gibson ökológiai észleléseleméletét integrálva (Gibson 1979) a testi elme gondolata az észlelést olyan kontextuális eseménysorként fogja fel, amely nem tétlen befogadóként, hanem exploratív operátorként éri tetten az észlelőt. Az elme, amely észlel, nem a *gondolkodó dolog* elszigeteltségében, hanem a testben való jelenlét aktivitásában működik. Ne hagyjuk félrevezetni magunkat a banális példamondatoktól („ez itt egy asztal”, „az ott egy fűszál”), amit észlelünk, az nagyon ritkán egy természeti fajta példánya, az észlelés nagyon ritkán egyszerű tárgyfelismerés. Például a legkevésbé sem közömbös, hogy mekkora az észlelő test az észlelt tárgy viszonylatában. Egy fűszál az embernek valami, amit átlép vagy letapos, a hangyának ellenben lehetséges útvonal, amit végigjár vagy nem (Johnson 2007). A sajáttest az észlelésben afféle kiterjedt elemeként funkcionál; mindent a saját léptéke szerint tekint, a környezetét a maga igényeinek megfelelően konkretizálja. Még az is kérdéses, hogy hol ér véget ez a test. Észlelő-feltáró tevékenységében a test integrálja azokat a kiegészítőket, amelyeket szokásosan használ (szemüveg, görbebot, művégtag, kéziszerszám). Miért tekintjük az elme részének az emlékezetet egy átlagosan működő személy esetében, és miért nem a jegyzetfüzetet olyan valakinél, aki elveszítette középtávú memóriáját, de a jegyzeteiből remekül eltájékozik a világban? Elvégre funkcionálisan és az eredmény szempontjából mindkettő ugyanoda vezet:



adatokra van szükség, ezeket rutinszerűen elő lehet hívni egy memóriatárból, és azonmód fel lehet őket használni (Clark–Chalmers 1998).

Az elmeként működő szituált test folytonos előre- és visszacsatolási viszonyban áll környezetével. Az exploratív észlelési szituáció nem egy kimerevített pillanat megjelenítése, hanem a viselkedés közvetlen jövőjének anticipációja és hatékonyságának korrekciója. Bateson példájával élve (Bateson 1972), a favágó minden egyes fejszezapása erejében és ívében igazodik a megelőző csapások által hasított formához; a favágás önkorrektíós folyamat, amelyben a test mozgása és a környezet változása egyetlen, dinamikusan módosuló rendszert képez. Ahogy Bateson leszögezi, ez az egész fa–szem–agy–izom–fejsze–sújtás–fa kör egy önszervező struktúra, amely „az immanens elme jellemzőivel rendelkezik” (Bateson 1972. 317).

Ez a hús-vér elme a maga szituált, szituációfeltáró és adottságmódosító dinamikájában, azaz a test ön maga számára is megjelenik. A testem mint afficiált/afficiáló ágens része a perceptív tapasztalatnak. Ám ez az önmegjelenés nem homogén; különbséget kell tenni testkép és testséma között (Gallagher 1995; 2007) – mindkettőnek megvan a maga szerepe a szituált testi tapasztalatban. A testkép nem más, mint a saját test intencionális tárgyként való megjelenése az önátélésben, amely tartalmazza a szubjektum perceptuális tapasztalatát önön testéről, konceptuális megértését e testről és emocionális attitűdjét a teste felé. A testkép úgy jelenik meg, mint ami a tapasztaló szubjektumhoz tartozik, irányulhat adott testrészeire vagy testrészekre, de az egész alkatra is. A testkép a szándékolt és tervszerű cselekvésben játszik szerepet, a tudatos figyelemben és a viselkedés önkontrolljában. Ugyanakkor a testkép hajlamos eltűnni a reflexió elől: a sztereotíp cselekvésekben felszívódik, egészen egyszerűen nem figyelünk rá. A testséma ellenben a test nem tudatos alanyisága mozgásai vonatkozásában; ő hajtja végre a cselekvéseket, amelyekkel a test mértéket vesz környezetéről és eligazodik benne. Szubperszonális és prenoétikus aktor. Még a tudatosan felügyelt, eltervezett mozgások esetében is a háttérben működik, hiszen minden gesztusunknak megvannak a tudattalan, készségszintű mozzanatai, kicsiny korrekciók a testtartásban, taglejtések, izomfeszítések. A testséma teljesítményei azok az apró, visszatérő megnyilatkozásformák, amelyek kinek-kinek egyedi megjelenési stílusát adják – egy-egy jellegzetes fejtartás, a gesztusok és pózok sorozata –, s amelyek anélkül egyénitenek, hogy szándékoltan kimunkáltak volnának. Testséma és testkép általában gördülékenyen együttműködik, a séma „alájátszik” a képnek, a kép besegít az új vagy bonyolult viselkedések begyakorlásában. Szétválásukról (és fogalmi elválasztásuk szükségességéről) bizonyos klinikai tapasztalatok révén szerezhetünk bizonyosságot. Híressé vált például Ian Waterman esete, aki egy idegrendszeri megbetegedés nyomán teljes egészében elveszítette proprioceptív tapasztalatát, vagyis a testséma tacit és megbízható működését, és a testképre támaszkodva kénytelen volt úgy megtanulni újra mozogni, hogy a legesekélyebb mozgását is aprólékosan előre el kellett ter-

veznie (Cole 1995). A testséma által biztosított nyugalom és önbizalom affektív értékei ezek felbolydulása vagy elvesztése után reflektálódnak. A már Merleau-Ponty által is tanulmányozott fantomvégtag-jelenség a testi öntapasztalás e két komponensének lehetséges összeütközésére világít rá. Amputált kezű vagy lábú személyek gyakran számolnak be elveszített végtagjaikban érzett fájdalomról vagy viszketésről, akár évekkel az operáció után is, amikor testképük már régen igazodott új életformájukhoz. Még érdekesebb, hogy az elvesztett kézben egyesek olyan göresöt éreznek, ami mintha valamilyen tárgy túlságosan erős megmarkolásából származna, és ettől az érzéstől semmilyen módon nem képesek megszabadulni. Amikor viszont egy olyan szerkezetbe helyezik ép karjukat, amely közepén egy tükörlapot tartalmaz, és így saját látószögükből úgy látják, mintha két egészséges karjuk volna (noha persze reflektáltan pontosan tudják, hogy erről szó sincs), akkor az ép kar ellazítása a fantomkar göresében is enyhülést okoz, legalábbis amíg a procedúra tart (Ramachandran–Blakeslee 1998). Itt egy szimulált testkép vesztegeti meg a testséma spontaneitását (mindenki hazudik – még a testképek is).

A testséma és a testkép megkülönböztetése fogalmi kulcsot adhat annak a furcsa fenomenális kettősségnek az értelmezéséhez is, ahogy a cselekvő én önmaga számára prezentálódik a testi tapasztalatban. Hiszen egyfelől kétségtelenül a szituációk sajáttest felőli koncentrátsága adja meg az alanyiség elemi evidenciáját (én vagyok ez a test, velem történik mindez, én iszom kávét tüstént). Másfelől azonban a testfenomén hajlamos a „fokális eltűnés” defenomenalizációjába hanyatlani, azaz a test mintegy kiveszi magát önnön tapasztalatából, a tárgyra és cselekvésre irányultság forgatagában az irányuló és cselekvő test visszahúzódik az állandó önmegjelenítésből, és testsémája csendjébe burkolózva tesz-vesz (mindaddig, amíg proprioceptív tapasztalata ép).

3. *Mozgásos konstitúció.* Szorosan a szituátság gondolatához kapcsolódik, hogy a környezet felmérése, a tárgyak és események azonosítása, a világ és a világ dolgainak kvalitatív megjelenése a testi szubjektum saját mozgási lehetőségeinek megfelelően bontakozik ki. Odalépek a dolgokhoz, lehajolok hozzájuk, arrébb teszem őket (ha bírom), változásaikat saját mozgásom függvényében értelmezem (így szabom meg, például, mi számít gyorsnak vagy lassúnak, nehéznek vagy könnyűnek, közelinek vagy távolinak). A mozgás révén szerzem meg az erő kifejtés és a térbeliség fundamentális tapasztalatait; előbbi a testemnek elengeszegülő tömegek és ezek elrendezésének változtathatósága vagy változtathatatlansága nyomán az okozás kategoriális absztrakciójához vezet, utóbbi pedig az irány, a pálya, az előrehaladás és a tartalmazás kulcsmetaforáival kognitív sémákat generál, amelyek a reflexió számára mindvégig (tudattalanul) irányadók maradnak (Johnson 2007). Voltaképpen a fogalmi gondolkodás terminusait szerzem ilyen módon meg, amely terminusok, noha a testi szubjektivitás öntapasztalatából származnak, mégis típusosak, így interszubjektíve működtethetőek lesznek. Valamennyiünk teste nagyjából egyforma, valamennyien nagyjából azonos

exploratív aktivitást fejtünk ki környezetünkben. A világba való értelmes behelyezkedés érdekében minden újszülöttnak meg kell tanulnia működtetni gesztusait, végtagjait, ujjait, hangképzését, hiszen mindez nélkülözhetetlen a kommunikáció, a tárgymanipulálás és a koordinált testmozgás szempontjából. Fontos, hogy mindezek nem nyelvi (részben nyelven kívüli, részben nyelv előtti) készségek, mégis az értelemkonstitúció feltételei.

A tárgykonstitúció is testléptékű, ám ez nem jelenti azt, hogy a testi elme elmélete valamiféle testfenomén-központú idealizmust építene fel, amely szerint az létezik, és úgy létezik, amit és ahogy a test megjelenít önmaga számára önnön alkati és önmegjelenítő képességei szerint. Az, hogy egy tárgy mi és miként jelenik meg, függ először is az élő test kifejlődő, önmozgató és önaffektív képességeitől, másodjára függ magától a tárgytól (hogy milyen szerkezetű, anyagú, állagú stb.), és függ a környezettől, a situációtól, amelyben előkerül (Gibson 1979). Az a nem éppen új keletű filozófiai belátás, hogy a tárgy nem elmefüggetlen, ekképp új tónust kap: nem elmefüggetlen, értsd: nem test- és mozgásfüggetlen.

4. *Affektivitás.* Az affektusok (érzelmek, indulatok, hangulatok) nem a funkcionális kognícióként felfogott problémamegoldó aktivitás merő kísérőjelenségei. A mindenkori situációk értékeléséhez (tehát a kompetens és releváns viselkedéshez) az affektusok nélkülözhetetlenek. Félelmet érzek egy nagy és ronda állat megjelenésekor, vonzalmat és nemi vágyat a potenciális partner láttán, unat, ha megint azt kell csinálnom, amit akkor szoktam csinálni, ha unatkozom, harci kedv támad bennem a vetélytárs jelenlétében, majd meghalok egy csésze kávéért – az élethelyzetek adta feladatok megoldásának az affektusok az elsőrendű motivációs tényezői. Nincs affektusmentes, azaz antikvalitatív helyzet; a közömbösség is csak érzelmi modalitás.

Az affektusok egyszerre kognitív, neurális és fenomenális testi folyamatok. Használjuk, működtetjük és érezzük őket, de ez utóbbi, az érzés, nem egy tetlen lélek intim birtoka, hanem azoknak a testi változásoknak a megtapasztalása, amelyek az affektusokat kinyilvánítják. A rémület esetében az izmok összehúzódnása, a mimika elmeredése, a gyomortáji feszülés, a szívdobogás, az egész testes lefagyás nem csupán a félelem velejárója, hanem *maga* a félelem mint érzés. Ez a fenomenális tapasztalat ízig-vérig a test tapasztalata (a kifejezés mindkét értelmében), és egyben annak tapasztalata, ahogyan a test reagál a helyzetére és módosítja azt. Úgy is lehetne fogalmazni, hogy az affektus a situáció egészét jellemzi (a félelem azt jelenti, hogy félelmetes helyzetben vagyunk); alkotója és része a situáció értelmének, és értelme végül is csak egy situációnak lehet. A testi elme teoretikusai előszeretettel hivatkoznak Dewey megfontolásaira, aki az affektív élményekben előálló fenomenológiai minőségekben, a kvalitatív tapasztalatban azt a szintézisreemtő elemet fedezte fel, amely mintegy átfogja és lehatárolja az egyes situációkat, elválasztja őket a többitől, és egyfajta koherenciát biztosít nekik (Dewey 1930/1988). Minthogy minden emberi

élethelyzet rendkívül komplex, heterogén – fizikai, társadalmi, kulturális, egyéni – összetevőkből álló struktúra, az, hogy egységes szituációként fenomenalizálódnak, korántsem triviális ügy. Az affektív kvalitás, amely velük jár, egyben egységteremtő erővel rendelkezik.

Az affektív szintézis a verbális kifejezésben is megnyilatkozik (Johnson 2007). Beszéd- vagy írásfolyamatainkat gyakran kíséri a siker vagy a sikertelenség érzése, az, hogy most végre fején találtuk a szöveget, vagy éppen az, hogy nem igazán azt mondtuk el, amit mondani szándékoztunk. A kifejezéskeresés – valamint az elégedettség vagy elégedetlenség – forrása a kvalitatíve egységes szituáció és a kifejezésmintázatok között *érezett* feszültség. A nyelvben készen álló panelek, sztereotíp fordulatok, vagyis a kész formák dolga az, hogy feldolgozzanak és kommunikálhatóvá tegyenek egy olyan helyzetet, ami még nem kész, nem zárult körülírható mintává – hiszen éppen benne vagyunk –, kvalitatíve azonban határozott arculattal bír. De a tapasztalati minőségek sokat hangoztatott kifejezhetetlensége elhamarkodott diagnózis. A megélt affektus és a kifejezési adekvátság elcsúszása adja a verbális közlés dinamikáját, stilisztikai és retorikai exploratív potenciálját. A kvalitatív tapasztalat nem szétfolyó, hanem nagyon is precíz: éppen ezért utasíthat el kifejezésformákat, ezért érezhetjük határozottan, hogy most valamit rosszul fejeztünk ki. A kijelentés az értelmet igyekszik megragadni; az értelem nem más, mint kvalitatíve egybefogott szituáció, a szituáció mindig egy test beágyazódása a világba – és ha az értelem testi, akkor a kijelentés értelme is testi eredetű (Johnson 2007).

A testi elme gondolata értelmében az, hogy történetesen éppen ilyen testtel rendelkezünk, és azt ilyen módon tapasztaljuk meg (és persze azért ilyen módon tapasztaljuk meg, mert ilyen testtel rendelkezünk), meghatározza mindazt, amit mentális tartalomnak, mentális állapotnak, tudatnak vagy elmének nevezünk. Sőt, a viszony valójában nem is a meghatározásé. A testi elme tétele szerint a test, a megjelenő test önmegjelenítési módjaiban kimeríti az elme fogalmát, annak minden tulajdonságával együtt. Még az elvont racionális gondolkodás is visszavezethető a fenomenális testtapasztalat kínálta kognitív metaforikára (Lakoff–Johnson 1999; Johnson 2007), mi több, a logikai formalizmusok sem képesek maguk mögött hagyni a testi tapasztalattal terhelt származásukat. Nincsen tehát test–elme probléma, egyedül test–test probléma van, nevezetesen az, hogy a fizikai mikrostruktúrákkal (is) jellemezhető és szaktudományosan kutatható test milyen viszonyban áll a magát holisztikusan és szituáltan megjelenítő fenomenális testtel. (Persze ez nem egyeduralkodó álláspont, Robert Hanna és Evan Thompson például egy elme–test–test problémát szerkeszt, amelyben a fenomenális test az, ami egyfelől mentális, másfelől fizikai aspektusokat hordoz (Hanna–Thompson 2003). A kvália kérdése azzal nyer megoldást, hogy mivel a fenomenalitás testi, a testtapasztalat pedig kvalitatív, a test egyszerű megjelenéséből következik, hogy *legyen neki valamilyennek lenni*, azaz hogy rendelkezzen minőségi élményekkel. A mentális okozás kizárási problémájára a testi elme eszméje azzal felel,

hogy mivel nincs miért azt gondolni, hogy az elme és a mentális tulajdonságok mások volnának, mint a test, ezért magától értetődő, hogy testi tulajdonságok és események oksági hatással lehetnek testi tulajdonságokra és eseményekre.

Nem biztos, hogy ennyire egyszerű a helyzet. Ám a további értékelés előtt tanácsos lesz röviden átgondolni azt, hogy a testi elme központi tétele, nevezetesen az, hogy amennyiben ilyen és ilyen testtel rendelkezünk (jobban mondva: ilyen és ilyen test *vagyunk*), akkor okvetlenül rendelkezniünk kell belső fenomenális tapasztalattal, megállja-e a helyét. Vagyis zombikra kell vadásznunk.

A filozófiai zombik olyan lények, amelyek fizikai és funkcionális felépítésüket tekintve a megtévesztésig olyanok, mint mi, csak éppen nincs fenomenális tapasztalatuk. Biológiai értelemben élnek, reagálnak a környezetükre, észlelnek, tanulnak, emlékeznek, vagyis kognitív működéseik rendben vannak, sőt, még arra is képesek, hogy veszekedjenek, szerelmet valljanak, és verseket írjanak, de mindehhez nem járul az a többlet, hogy nekik ez az egész *valamilyen* volna, kvalitatíve üresek.

A zombik témáját eredetileg Robert Kirk dobta be a köztudatba a hetvenes években (Kirk 1974/a; 1974/b), később David Chalmers elevenítette fel az ötletet (Chalmers 1996). A gondolat kísérlet célja egy antifizikalista érv kidolgozása: amennyiben zombik lehetségesek, a fizikai és funkcionális felépítés nem hordoz (okoz) szükségszerűen tudatos mentális tapasztalatot, következésképpen a tudatos tapasztalat valami plusz minőséget képvisel az agyi-idegi-kognitív infrastruktúrához képest, magyarázatához tehát, ahogyan Chalmers mondja, valamilyen „extra tartozék” kívánatik meg. Egyébként Chalmers előadásában az érv lefegyverzően egyszerű – ha valami elgondolható, akkor lehetséges, a zombik elgondolhatóak, tehát a zombik lehetségesek, ha pedig a zombik lehetségesek, a fizikalizmus bukott elmélet. Hozzá kell tenni, hogy ez az egész jóval több merő teoretikus tréfánál; Chalmers jórészt erre az évrre alapozza a maga naturalista tulajdonságdualizmusának tézisének, amelyben a tudatos tapasztalat a világ olyan alapvető jellemzője, amilyen a tömeg, a töltés vagy a téridő, ám nem fizikai jellemző. Ennél most fontosabb azonban, hogy a zombik lehetősége a testi elme gondolatával szemben is ellenérvként működik. Hiszen ha egyszer zombik lehetségesek, akkor lehetséges az, hogy ilyen és ilyen testi felépítés és szituáltság nem implikál óhatatlanul fenomenális öntapasztalatot, márpedig, ismétlem, ez a testi elme tézisének központi magva.

A zombik lehetőségét sokan és sokféleképpen vitatták. Robert Kirk, aki időközben feladta saját korábbi álláspontját a zombik lehetőségességéről, és szenvedélyes zombi-ellenes harcosná vált, egyik újabb könyvében (Kirk 2005) összefoglalja a zombivilág lehetetlenségével kapcsolatos megfontolásokat, és maga is előterjeszt egy ellenérvet. Nézzünk egy hozzávetőleges listát:

1. Chalmers helytelenül használja a modális fogalmakat. Abból, hogy a zombik elgondolhatóak, még nem következik, hogy lehetségesek, mert egy-

részt az elgondolhatóság nem implikál lehetségeséget, vagy ha igen, akkor is legfeljebb *logikai* lehetségeséget, nem pedig *metafizikai*t, holott itt erről volna szó.

Messzire vezetne, ha most beleártanánk magunkat a modális szemantikába, ezért inkább csak jelzem, hogy ezt az ellenvetést maga Chalmers is felveti könyvében, és saját kétdimenziós szemantikája segítségével elhárítja.

2. A fenomenális tapasztalat nem olyan, mint valami zakó, amit az ember ledob vagy elhagy, de attól még ugyanaz marad. A fenomenális tapasztalat az ember lényegi jellemzője, nélküle nem lenne az, ami, következésképpen a zombi nem is lehetne pontosan ugyanolyan, mint egy ember.

Ez nem jó érv, hiszen azt ugyan megtehetjük, hogy az ember definíciójába beemeljük a fenomenális tudatosságot, illetve hogy azt megtesszük lényegi tulajdonságának, de nem az a lényeg, hogy a zombik emberek-e, hanem hogy összetéveszhetőek-e az emberekkel. Nem emberek, nem is *pontosan* ugyanolyanok – ha pontosan ugyanolyanok volnának, akkor emberek volnának, nem pedig zombik –, hanem olyanok, *mintha*. És ez elég.

3. A zombi-hipotézis megkívánja a belső szubjektív tapasztalat teljes epifenomenalitását. Ha ugyanis a tapasztalat nem volna epifenomenális, akkor valamiképpen módosítaná a fizikai világot, ha pedig módosítaná, akkor ezen módosulatok hiányában a zombik nem lehetnének a megszólalásig hasonlóak hozzánk. Csakhogy ha zombikról beszélünk, akkor a kvalitatív tapasztalat hiányáról beszélünk, azaz – tagadólag – a kvalitatív tapasztalatra referálunk. Márpedig a referencia oksági felfogásából következően csak olyasmire tudunk sikeresen referálni, amivel oksági kapcsolatban állunk. Tehát ha sikeresen referálunk a kvalitatív tapasztalatra, akkor az nem lehet okságilag hatástalan, nem lehet merőben epifenomenális, és ha a kvalitatív tapasztalat nem epifenomenális, hanem okságilag hatékony, akkor a zombik elgondolhatatlanok. Szóval, ha elgondoljuk őket, nem gondolhatjuk el őket.

Ez nagyon szép formájú érv, csak sajnos megköveteli a sikeres referálás oksági elméletének szigorú elfogadását. Amennyiben ezt nem tesszük (mert például az első fejezetnél tovább olvassuk Putnam *Reason, Truth, and History* című könyvét), az ellenérv elveszíti erejét.

4. Kirk saját érve: A zombivilág tehát a kvalitatív tapasztalatra vonatkozó epifenomenalizmust implikál. Mármost az emberek nemcsak abban különböznek a zombiktól, hogy rendelkeznek kvalitatív tapasztalatokkal, de abban is, hogy ezeket a tapasztalataikat képesek észrevenni, felfigyelni rájuk, gondolkodnak rajtuk, összevetik őket egymással (ki tagadná, hogy naphosszat ezt tesszük?). A figyelem, a gondolkodás, az összetevés kognitív funkciók, amelyek részei az oksági univerzumnak. Ám oksági funkciókat csak olyasmire képes beindítani, ami maga is rendelkezik oksági kapacitással, márpedig ha a fenomenális tapasztalat rendelkezik vele, akkor

nem epifenomenális – és ekkor, egy egyszerű *modus tollens* segítségével oda jutunk, hogy a zombivilág ki van zárva.

Ez az ellenvetés azon alapul, hogy a kvalitatív tapasztalatok észrevétele, végiggondolása és egymással való összevetése valami *más* dolog, mint maga a tapasztalat. Azonban ez fenomenológiailag egyáltalán nem magától értetődő. Külön mentális aktus volna az, hogy élvezem a kávé ízét, és hogy észreveszem, hogy élvezem a kávé ízét? Aligha. A tapasztalatoknak ezzel a kognitív arzenállal való megkettőzése hiteltelenné teszi Kirk érvének intuitív alapját, és ez az infláció az érv egységét elértekteleníti.

Megközelíthető a kérdés egyszerűbben is. Ha elfogadjuk a zombik lehetőségét, akkor ezzel azt is elfogadjuk, hogy gyakorlatilag megkülönböztethetetlenek tőlünk – elvégre ezt már a zombi fogalma is tartalmazza. Amennyiben ténylegesen ez a helyzet, akkor nem lehetséges olyan eljárás, amellyel pusztán külsődleges megfigyelés – tehát nem valamiféle misztikus beleélés (bele? mibe?) – útján meg lehetne mondani, hogy egy adott helyen egybegyűlt, embernek tűnő lények közül melyik zombi és melyik nem. Röviden: nem lehetséges zombi-teszt. Amennyiben ugyanis elvileg lehetséges ilyesmi, akkor máris elesett a zombik legfőbb ismérve, hogy tudniillik külsőleg és funkcionálisan összetéveszthetők velünk. Ha van zombi-teszt, akkor zombik nincsenek, és nem is lehetségesek (és nincs szükség a tesztre). Márpedig egy efféle teszt könnyedén megkonstruálható: adjunk oda a jelenlevőknek, mondjuk, két-két pohár teljesen egyforma hatóanyagú, de eltérő ízesítésű köhögés elleni szirupot, egy citromízűt és egy málnaízűt, és mondjuk nekik azt, hogy igyák meg azt, amelyik jobban ízlik nekik. Egy ilyen döntést nyilván csakis kvalitatív tapasztalás függvényében lehet meghozni. Aki belekóstol ebbe, majd abba, azután kiüssza az egyiket, ember, aki csak kortyolgat és áll tanáctalantul, az zombi. Persze erre bárki mondhatná, hogy egyrészt a zombi képes lehet arra, hogy véletlenszerűen válasszon a két lehetőség közül, és az esetleges ellenőrző kérdésre („ízlett?”) válaszolhatná azt, hogy „fincsi volt” – habár hogy egy zombi miért hazudna (úgy értem, mi volna abban a *jó* neki), külön kérdés –, másrészt pedig az emberi lények is gyakran annyira elesettek az ízlésbeli ügyekben. Jó, a teszt kiterjeszthető. Ne egy feladat legyen, hanem olyan feladatsor, amelynek minden eleme valamilyen ízlésre, érzésre, minőségre vonatkozó választást irányoz elő, és amely alkalmas arra, hogy egy ízlésprofil rajzoljon fel. Mindannyian kissé kapkodók és szeszélyesek vagyunk ugyan, de ízekben, illatokban, színekben, arcvonásokban valamennyien tesztelhetők vagyunk vonzalmaink szerint, és ezek a vonzalmak nem teljesen inkoherensek. Márpedig vonzalmainkat igenis belső, szubjektív fenomenális tapasztalataink realizálják, és mivel a zombinak ilyen nincs, lebukik. Még akkor is, ha valamilyen evolúciós funkció a megtévesztést sugallja neki, mert még hazudni sem képes koherens ízlésprofil, hiszen annak koherenciája nem értelmezhető a kvalitatív tapasztalaton kívül. A zombi soha nem volna képes eldönteni, hogy

Bach vagy Beethoven, Fichte vagy Schelling, Chandler vagy Hammett, Angelina vagy Scarlett, hamburger vagy hot-dog. A zombi lebukik, ami pedig azt jelenti, hogy nem is létezett soha, mert az igazi zombi nem tesztelhető. Zombik nem léteznek, a priori tehát nincs kizárva, hogy a testi elme gondolata érvényes legyen, az élőhalottak éjszakája véget ért.

A testi elme teóriája az elmefilozófiai rejtélyt egy kétaspektusos, erősen emergens fogalmiságban véli megoldani. A kétaspektusosság azonosságelmélet, de különbözik annak hagyományos alternatíváitól. Nem típusazonosság, hiszen nem azt állítja, hogy az egyes testi-mentális állapotok típusai volnának azonosak egyes fizikai állapotok típusaival, hanem azt, hogy a hordozó (a test) ugyanaz, és ennek vannak eltérő – fizikai és fenomenális – aspektusai. Nem is példányazonosság, mert nem azt állítja, hogy egyedi fenomenális állapotokat egyedi fizikai állapotok realizálnak, hanem azt, hogy kettős realizálás megy végbe: a fizikai állapotok ténylegesen megvalósítanak fenomenális állapotokat, de visszafelé is igaz ez, a fenomenális állapotok megvalósítanak fizikai állapotokat. Sejtek, szövetek, kémiai reakciók nélkül nem létezhetne semmi olyasmi, ami a világban megjelenik, ugyanakkor azt, hogy a világ olyan, amilyennek megjelenik, hogy a fizikai entitások éppen olyan struktúrákba szerveződnek, amilyenekbe, a fenomenális szféra valósítja meg a maga exploratív aktivitásával. Ez a kétaspektusosság az erős emergencia magyarázati kategóriájával egészül ki. Az erős emergencia azt jelenti, hogy egy bizonyos szerveződési komplexitás szintje fölött egyszerű elemek vagy szerkezetek olyan új jelenségesoportot hoznak létre, amelynek működése elyben sem vezethető le az alkotóelemek működéséből. A testi fenomenalitás ilyen komplex és autonóm működéssel bíró jelenség volna, amely a fizikai entitások és események magasan szervezett összekapcsolódásából állna elő.

Csakhogy az emergencia fogalma oksági aszimmetriát hordoz. Az alacsonyabb szintű jelenségek viselkedése okságilag felelős a magasabb szint megjelenéséért (ha tulajdonságaiért közvetlenül nem is), de a magasabb szintű folyamatok önmagukban nem képesek oksági hatást előidézni az alacsonyabbakon – ez a lefelé okozás problémája. A lefelé okozást pedig éppen az oksági zártság elve miatt kell problematikusnak minősíteni. Ha az alacsonyabb szintű folyamatok okságilag nem zártak, azaz viselkedésük teljes magyarázata nem alkotható meg a saját szintjükön, akkor azt kellene feltételezni, hogy oksági nyitottságuk valamilyen előkészület a magasabb szint hatásainak befogadására – ez pedig a legcsúfosabb teleológiához vezet. Konkretizálva ezt a fenomenális tapasztalat és a fizikai mikrostruktúrák esetére, ha a fizikai entitások és események viselkedésébe az emergens fenomenális események beavatkozhatnának oly módon, hogy ez a beavatkozás nem volna visszavezethető magukra a tisztán fizikai jelenségekre, akkor oksági rést kellene feltételeznünk a fizikai világban, még hozzá specifikus oksági rést, amely éppen a fenomenálisan öntapasztaló makrotest hatásaira van szabva. Ez pedig kevésbé plauzibilis feltételezés. Az emergencia magyarázó ka-



tegoriáját tehát – bármennyire kézenfekvőnek mutatkozik is első látásra – alaposan felül kell vizsgálni.

Ám ugyanez a megfontolás a kétaspektusosság tanát is fenyegeti. Nem elég annyit mondani, hogy a fizikai mikrostruktúra a világ *egyik* aspektusa, a fenomenológiai makrostruktúra pedig a *másik*. Hiszen kénytelenek vagyunk elfogadni, hogy a fizikai mikrostruktúra viselkedése okozza azt, hogy a makrostruktúra (a testléptékű, szituatív konstelláció) éppen olyan, amilyen, fordítva viszont ez sem áll. De pontosan itt lehet hasznosítani a kettős realizálás elvét. Amikor azt mondjuk, hogy a fizikai mikrostruktúra viselkedése okozza, hogy a makrostruktúra ilyen és ilyen, ebbe beleértjük, hogy ez a bizonyos makrostruktúra, a dolgok világa, hogy egyszerűbb legyen a képlet, valami entitás vagy állapot, ami önmagában fennáll. Mikrostrukturális szempontból (ha volna ilyen szempont, és ez a lényeg) azonban a makrodolog (test, szituáció) értelmezhetetlen. Nem áll elő, nem válik le szintén mikrostrukturális környezetéről. A szigorúan vett oksági folyamatok nem testléptékű közegben zajlanak. A fizikai magyarázat, az egész fizikai és fizikalista megközelítés eleve a fenomenális szintről indul, abba ágyazódik bele, annak miérteit kutatja. Ez nemcsak abban a banális értelemben igaz, hogy ha nem volna fenomenális tapasztalat, akkor semmiféle kutatás nem volna, hanem abban is, hogy az okság kategóriája kiküszöbölhetetlenül a testléptékű jelenségek körében ragad.

Akkor tehát a testi elme gondolata megoldotta az elmefilozófiai rejtélyt? Azt állítom, hogy nem, de azt is állítom, hogy beállítottságának módja és konkrét kutatási eredményei megfelelő *nyersanyagot* szállíthatnak ahhoz, hogy a megoldáshoz nekiláthassunk. Alkalmas lehet arra, hogy a reprezentációról, pozicionális attitűdökről, intencionalitásról szóló beszédmód helyett módot adjon e fogalmi háló átrendezésére – a test–test probléma végiggondolása kulcsmozanat lehet. Nyersanyagról és beállítottságmódról beszélek, mert a testi elme mindenekelőtt egy kutatási program, amelynek eredményei hármassal valószínűsítésű procedúrán méretnek le. Megállapításainak először is fenomenológiai értelemben genuin tapasztalatot kell megragadniuk, vagyis találkozniuk kell a tapasztalatra vonatkozó intuíció evidenciájával. Azután e megállapításokból következő hipotéziseket a kognitív viselkedéskutatás kísérleti eszközeivel kell ellenőrizni, vagy klinikai esetleírásokkal kell visszaigazolni. Végül képalkotó eljárások (vagy más, a fiziológiai folyamatokat monitorozó technikák, például számítógépes modellalkotás) segítségével fel kell tárni a megállapításokban jellemzett tapasztalat és viselkedés strukturális hátterét. Bármelyik marad is el ezek közül, a megállapítások érvényessége nem számít biztosítottnak.

Nem feladatokat osztok, hanem egy ígéretet értékelek. És ha jobban megfontoljuk, a huszadik század negyvenes–ötvenes éveiben Merleau-Ponty tényleg valami nagyon hasonlóan látta a testtel kapcsolatos filozófia esélyeit.

## IRODALOM

- Anderson, Michael L. 2003. Embodied Cognition: A Field Guide. *Artificial Intelligence*. 149/1. 91–130.
- Bateson, Gregory 1972. *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution and Epistemology*. Chicago, University of Chicago Press.
- Bermúdez, José Luis – Antony Marcel – Naomi Eilan (szerk.) 1995. *The Body and the Self*. Cambridge/MA, MIT Press.
- Brooks, Rodney A. 1991. Intelligence without Representation. *Artificial Intelligence*. 47/1–2. 139–159.
- Brooks, Rodney A. 1999. *Cambrian Intelligence: The Early History of the New AI*. Cambridge/MA, MIT Press.
- Carman, Taylor 1999. The Body in Husserl and Merleau-Ponty. *Philosophical Topics*. 27/2. 205–226.
- Chalmers, David J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford, Oxford University Press.
- Churchland, Paul M. 1981. Eliminative Materialism and the Propositional Attitudes. *Journal of Philosophy*. 78/2. 67–90.
- Clark, Andy 1997. *Being There: Putting Brain, Body and World Together Again*. Cambridge/MA, MIT Press.
- Clark, Andy – David J. Chalmers 1998. The Extended Mind. *Analysis*. 58/1. 7–19.
- Cole, Jonathan 1995. *Pride and a Daily Marathon*. Cambridge/MA, MIT Press.
- Dewey, John 1930/1988. Qualitative Thought In *uő The Later Works, 1925–1953* (5. köt., 1929–1930). Carbondale/IL, Southern Illinois University Press. 243–263.
- Dreyfus, Hubert L. 1992. *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge/MA, MIT Press.
- Dreyfus, Hubert L. 2002. Intelligence without Representation: Merleau-Ponty's Critique of Mental Representation. *Phenomenology and Cognitive Sciences*. 1/4. 367–383.
- Ellis, Ralph 2000. Efferent Brain Processes and the Enactive Approach to Consciousness. *Journal of Consciousness Studies*. 7/4. 40–50.
- Gallagher, Shaun 1995. Body Schema and Intentionality. In José Luis Bermúdez – Antony Marcel – Naomi Eilan (szerk.) 1995. *The Body and the Self*. Cambridge/MA, MIT Press, 225–244.
- Gallagher, Shaun 2005. *How the Body Shapes the Mind?* New York, Oxford University Press.
- Gallagher, Shaun 2007. Phenomenological and Experimental Contributions to Understanding Embodied Experience. In Tom Ziemke – Jordan Zlatev – Roslyn M. Frank (szerk.) *Body, Language and Mind. Volume 1: Embodiment*. Berlin, Mouton de Gruyter. 271–294.
- Gelder, Timothy van. 2005. Beyond the Mind–Body Problem. In Christina E. Erneling – David Martel Johnson (szerk.) *The Mind as a Scientific Object: Between Brain and Culture*. Oxford, Oxford University Press. 457–470.
- Gibbs, Raymond W. Jr. 2005. *Embodiment and Cognitive Science*. Cambridge, Cambridge University Press.
- Gibson, James J. 1979. *The Ecological Approach to Visual Perception*. Boston, Houghton Mifflin.
- Hanna, Robert – Thompson, Evan 2003. The Mind-Body-Body Problem. *Theoria et Historia Scientiarum: International Journal for Interdisciplinary Studies*. 7/1. 23–42.
- Husserl, Edmund 1989. *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy, Second Book: Studies in the Phenomenology of Constitution*. Ford. Richard Rojcewicz, André Schuwer. Dordrecht, Kluwer Academic Publishers.
- Jackson, Frank 1982. Epiphenomenal Qualia. *Philosophical Quarterly*. 32. évf. 127. sz. 127–136.

- Johnson, Mark 2007. *The Meaning of the Body. Aesthetics of Human Understanding*. Chicago, University of Chicago Press.
- Kampis Gy. 2002. A gondolkodó test. *Magyar Tudomány*. 109 (új folyam 47)/1. 33–40.
- Kampis Gy. 2002. Közelíthető-e egymáshoz a test és a tudat? In Vizi E. Szilveszter – Altrichter Ferenc – Nyíri Kristóf – Pléh Csaba (szerk.) *Agy és tudat*. Budapest, Books in Print. 106–118.
- Kim, Jaegwon 1993. The Myth of Nonreductive Materialism. In uő *Supervenience and Mind: Selected Philosophical Essays*. Cambridge, Cambridge University Press. 265–284.
- Kim, Jaegwon 2005. *Physicalism, Or Something Near Enough*. Princeton, Princeton University Press.
- Kim, Jaegwon 1993/2008. A non-reduktivista gondjai a mentális okozással. Ford. Eszes Boldizsár. In Ambrus Gergely – Demeter Tamás – Forrai Gábor – Tözsér János (szerk.) *Élméofilozófia*. Budapest, L'Harmattan. 112–134.
- Kirk, Robert 1974/a. Sentience and Behaviour. *Mind*. 83. évf. 329. sz. 43–60.
- Kirk, Robert 1974/b. Zombies v. Materialists. *Proceedings of the Aristotelian Society*. Supplementary vol. 48. 135–152.
- Kirk, Robert 2005. *Zombies and Consciousness*. Oxford, Clarendon Press.
- Kripke, Saul 1972/2007. *Megnevezés és szükségszerűség*. Szerk., bevez. Zvolenszky Zsófia; ford. Bárány Tibor. Budapest, Akadémiai Kiadó.
- Lakoff, George – Mark Johnson 1999. *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York, Basic Books.
- Levine, Joseph 1983. Materialism and Qualia: The Explanatory Gap. *Pacific Philosophy Quarterly*. 64. 354–361.
- McGinn, C. 1989/1999. Megoldható-e a test-lélek dualitásának rejtélye? Bevez., ford. Reke Lajos. *Gond*. 20. 185–204.
- Merleau-Ponty, Maurice 1942. *La structure du comportement*. Paris, Presses Universitaires de France.
- Merleau-Ponty, Maurice 1945. *Phénoménologie de la perception*. Paris, Gallimard.
- Nagel, Thomas 1974/2004. Milyen lehet denevérenek lenni? Ford. Sutyák Tibor. *Vulgo*. 5/2. 3–12.
- Putnam, Hilary 1975. The Nature of Mental States. In uő *Mind, Language and Reality. Philosophical Papers Volume 2*. Cambridge, Cambridge University Press. 429–440.
- Putnam, Hilary 1975/2010. A 'jelentés' jelentése. Ford. Kovács János és Polgárdi Ákos. *Különbség*. 10/1. 13–74.
- Ramachandran, Vilayanur S. – Sandra Blakeslee 1998. *Phantoms in the Brain*. London, Fourth Estate.
- Seager, William 1995/2004. Tudat, információ és pánszichizmus. Ford. Sutyák Tibor. *Vulgo*. 5/2. 47–62.
- Smith, David Woodruff 1995. Mind and Body. In Barry Smith – David Woodruff Smith (szerk.) *The Cambridge Companion to Husserl*. Cambridge, Cambridge University Press. 323–393.
- Szigeti Attila 2002. A tudatosság: rejtély vagy képesség? Az analitikus filozófia és a kognitív tudomány találkozása a fenomenológiával. *Kellék*. 22. 81–122.
- Varela, Francisco J. – Evan Thompson – Eleanor Rosch 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge/MA, MIT Press.

## Transzcendentális és testiség a husserli fenomenológiában

Es muss doch möglich sein,  
Faktum und ideale Möglichkeit und Notwendigkeit  
in ein wissenschaftliches Verhältnis zu setzen!  
(Husserl)

### I. A PERFEKT TUDAT

A mottónak választott idézetnél aligha lehetne tömörebben és elszántabban összefoglalni, mire alapozta Husserl a transzcendentális fenomenológiát mint univerzális tudományt: „mégiscsak lehetségesnek kell lennie annak, hogy tudományos viszonyba állítsunk faktumot és ideális lehetőséget és szükségszerűséget” (Husserl 1973. 131). Persze, kérdezhetnénk, mi mást nevezünk tudománynak, ha nem a megismerésnek azt a módját, amelynek során az esetleges tényekből szükségszerű törvényekhez jutunk? Csakhogy az idézet nem ezt mondja. Nem azt jelenti, amire elsősre gondolunk, hogy tudniillik a tudományoknak az a dolguk, hogy indukciót alkalmazva jussanak el az esetleges, egyedi tényekből a tényeket előidéző és kormányzó általános és szükségszerű törvényekhez. Hanem azt, hogy lehetségesnek *kell* lennie, hogy a faktumot közvetlenül, minden indukció és absztrakció nélkül, mint ideális lehetőséget és szükségszerűséget felmutassuk. Mert minden faktum szükségszerű. De minek kapcsán merül fel a legélesebben esetlegesség és szükségszerűség problémája a transzcendentális fenomenológiában? A legalapvetőbbben és a legnagyobb téttel bíró módon természetesen a tudat kapcsán.

Hogy kontextualizáljuk a mottót, induljunk ki két idézetből, amelyeknek a tartalma látszólag ellentmondásban áll egymással. Az egyik *A fenomenológia ideája* címen közzétett korai szövegből való, amely a Husserl által 1907-ben Göttingenben tartott előadássorozat publikált változata. Az itt közölt első előadásban Husserl a megismerés lehetőségének problémáját veti fel. A természetes gon-

dolgodás számára, mondja, a megismerés lehetősége magától értetődő, nem csupán kétségtelen, de kérdésre sem érdemes tény. Mihelyt azonban reflektálni kezdünk az ismeret és a tárgy viszonyára, vagyis a természetes gondolkodás talajáról átlépünk a filozófiára, minden problematikussá válik. A problematikusság ebben az esetben azt jelenti, hogy mivel a tárgy csak szubjektív adottságmódok összességeként lesz a megismerés tárgya, vagyis relatívvá válik, a megismerés és az ismeret elveszíti az őket kitüntetett módon jellemző szükségszerűséget. A tárgy relativitását csak az szüntetheti meg, ha a szubjektivitás, melynek minden objektivitás kétségtelenül a függvénye, maga szükségszerűnek bizonyulna. Ebben az esetben a tárgy továbbra is relatív maradna a tudathoz való viszonyában, azonban a róla származó ismeret, minthogy az abszolút(ként tekintett) tudatban jön létre, már abszolút érvényű lenne. Csakhogy, és itt következik az említett első idézet, a szubjektív pólussal, tehát a logikus gondolkodással kapcsolatban is felmerülhet a kétely, hogy nem abszolút szükségszerű, hanem véletlenszerűen esetleges:

Biológiai gondolatok nyomulnak előtérbe. Emlékeztetnek bennünket a modern fejlődélméletre, mely szerint az ember a létért való küzdelemben és a természetes kiválasztódás során fejlődött, s vele együtt fejlődött az értelme is, s értelmével együtt az azzal velejáró s azt jellemző formák, így közelebbről a logikai formák is. Eszerint a logikai formák és logikai törvények csupán az emberi faj véletlenszülte mivoltát fejezik ki, és ez másként is alakulhatott volna, s a jövőbeni fejlődés folyamán mássá is lesz? A megismerés tehát *emberi megismerés* csupán, s *az emberi intellektuális formákhoz* kötve képtelen arra, hogy a dolgok természetét, magukat a dolgokat magánvalóságukban „eltalálja” (Husserl 1972. 48).

Bár a kérdező forma miatt nem tűnik állásfoglalásnak, mégis – ismerve a kérdező forma filozófiai szerepét – határozottan állíthatjuk, hogy Husserl, Nietzschevel ellentétben, és talán ellenében is, egyértelműen elveti ennek az antroponológiai fejlődésszémének a relevanciáját, legalábbis a megismerés tekintetében. A logika, a logikus gondolkodás *nem eredhet* a logikátlanságból, a megismerés és az igazság nem lehetnek a létfenntartás sikeresnek bizonyult kellékei, és az emberi megismerés nem csupán az egyik lehetséges – az emberi intellektuális formák által meghatározott – formája a megismerésnek.<sup>1</sup> A fenomenológia mindenki számára ismert kiindulópontja és tartós alapja az a meggyőződés lesz, hogy az emberi tudat nem esetleges, hanem abszolút szükségszerű faktum, ennek a tudatnak a kifejlődése nem alakulhatott volna másképpen, és semmilyen jövőbeni fejlődés nem változtathatja meg. Tudjuk, hogy ennek a meggyőződésnek a háttérében a pszichologizmus-kritika áll, amelynek célja a logika, a logikai tör-

<sup>1</sup> Vö. Nietzsche 1997. 110, 111. aforizmak.

vények, műveletek és tárgyak önállóságának, a gondolkodás pszichofizikai aktu-saitól való függetlenségének kimutatása és megőrzése, annak garantálása, hogy a logikai ítéletek az ítélés esetleges pszichofizikai folyamatától független, örök igazságok. A logikában megállapított szükségszerűséget Husserl a tudati tevékenység egészére kiterjeszti. Akár úgy is értelmezhetjük a fenomenológia kibontakozásának történetét, hogy az úgynevezett transzcendentális fordulat ennek a kiterjesztésnek a következménye vagy motívuma, amely kiteljesedett formáját az eidetikus redukción alapuló lényegelemzésekben éri el. A noétikus lényegelemzések hivatottak kimutatni, hogyan és miért éppen így épül fel szükségképpen a transzcendentális szubjektivitás és interszubjektivitás teljesítményeként a világ.

Az interszubjektivitás azonban az eltérése és a pontosításon keresztül elkerülhetetlenül a változás és alakulás, egyszóval a kibontakozás értékét írja bele a megismerés problematikájába. Másképpen: történetivé teszi, ha nem is az igazságot, de a megismerést mindenképpen. Az ismeretek vagy az igazság története egyben a tudatnak a története is, melynek alapja maguknak az egyedi tudatoknak a története, ez pedig, *per definitionem*, máris függetleníthetetlen az egyedek biológiai adottságaitól. Ez pedig felveti azt a kérdést, hogy nem kell-e a tudat genezise kapcsán mégiscsak elfogadnunk valamilyen fejlődéskoncepciót, amely nagyon közelinek tűnik a biológiai evolúció gondolatához. Következzék most a második ígért idézet másfél évtizeddel későbből, a húszas évek elején tervbe vett szisztematikus műből:

A világ csak akkor létezhet, ha konstitutív módon kifejlődik, ha az abszolút szubjektivitás kifejlődik, ha maga a világ kifejlődik úgy, hogy az emberi öntudat formájában öntudattá, majd még tovább tudományos öntudattá fejlődik (Husserl 1973. 136).<sup>2</sup>

Mit gondoljunk erről a kétféle fejlődésmérőről: a biológiai evolucionizmusról, amelynek elfogadása a tudat és a logika esetlegességéhez vezetne, és amelyet ezért el kell vetni, illetve a világ és az öntudat kifejlődéséről, melyben minden szükségszerűnek bizonyul, és amely így létjogosultságot nyerhet a transzcendentális fenomenológiában? Első megközelítésben azt mondhatjuk, hogy a két idézet két különböző korszakból, de mindenképpen két különböző kontextusból való: az egyik a statikus fenomenológia, a másik a genetikus fenomenológia korszakából vagy kontextusából. Husserlben a statikus vizsgálódásokról a genetikus vizsgálódásokra való áttérés motívumaként megszületett egy felismerés, melynek eredményeképpen már nemcsak lehetségesnek, de egynesen szük-

<sup>2</sup> „[D]ie Welt kann nur sein, wenn sie sich konstitutiv entwickelt, wenn die absolute Subjektivität sich entwickelt, wenn sie selbst die Welt sich entwickelt, so dass sie sich zum Selbstbewusstsein in Form von menschlichem entwickelt, und sich dann weiter zum wissenschaftlichen Selbstbewusstsein entwickelt.”

ségesnek is tartotta, hogy a tudat genetikus történetéről beszéljen, ám ez nem járt azzal, hogy a tudatot kiszolgáltatva volna az esetlegességnek.

Nem tagadva azt, hogy ennek az irányváltásnak lényegi szerepe van a fenomenológia kibontakozásában, és hatása van a fenomenológia alapelveire és alapfogalmainak értelmére is, beláthatjuk, hogy az előbbi koncepciót elvetni nem jelenti azt, hogy az utóbbit ne lehetne elfogadni, hiszen a két idézetben nem ugyanolyan értelemben vett fejlődésről esik szó. Bármennyire közelinek látszik is a genetikus fenomenológiában elfogadott fejlődéseszmé a biológiai evolucionizmushoz, a két fejlődéskoncepció lényegi ponton különbözik egymástól. A biológiai evolucionizmus szükségszerű elv alapján (a létért folytatott küzdelem nevében), de összességében végecéll nélkül zajló, esetleges folyamatként mutatja be a fejlődést, a tudat genetikus története viszont olyan teleologikus fejlődés, amely esetlegességeken (egyedi tudatok kifejlődésén) keresztül halad a szükségszerű cél felé. A különbség tehát a fejlődés célszerűségében, illetve ennek hiányában áll. Vagyis az esetlegesség és szükségszerűség viszonyára irányuló kérdésünk kiegészül a célszerűségre irányuló kérdéssel, vagy azzá alakul át. Már nem (csak) azt kell kérdeznünk, hogy esetleges, vagy szükségszerű-e a tudat, hanem azt is, hogy mi a célja a tudat genetikus történetének.

Ahol van cél, ott a célhoz vezető út egyes állomásai csak látszólag esetlegesek. Ezért kell lehetségesnek lennie, hogy a faktumot mint ideális lehetőséget és szükségszerűséget ragadjuk meg. Az *Eszmék* II. részében, a szellemi világ konstitúciójának elemzése során találjuk az alábbi észrevételt, melyre a későbbiekben még visszatérünk:

Ha a közösségi szubjektumok az individuális vagy interindividuális módon megjelenő tárgyak esetében a jelenségről egybehangzó ítéletet (*Erscheinungsprädikate*) alkotnak, az csupán véletlenszerű faktum, ha ellenben a tapasztalatot feldolgozó ésszerű gondolkodás (*vernünftiges Erfahrungsdenken*) során az objektív (fizikai) meghatározottságok tekintetében jutnak meggyezésre, az szükségszerű (Husserl 1984. 38).<sup>3</sup>

Az interszjektív tudati aktusok ugyanis nemesak hogy nem az egyén individuális pszichofizikai adottságaihoz, de nem is esetlegesen az emberi léthez kötődnek, hanem az egyetemes tudat tiszta lényegiségei, amelyek totális szervezetségükben az ész(szerűséget) mint olyat alkotják. Ahogy a *Kartezianus elmélekedésekben* olvassuk: „Az ész nem esetleges-faktikus képesség, nem lehetségszerűen esetleges tényállások címszava, hanem általában a transzcendentális

<sup>3</sup> „Jede Übereinstimmung der Gemeinschaftssubjekte hinsichtlich der Erscheinungsprädikate der individuell und interindividuell erscheinenden Objekte gilt als zufälliges Faktum, dagegen Übereinstimmung hinsichtlich aller im vernünftigen Erfahrungsdenken sich herausstellenden objektiven (physikalischen) Bestimmungen als notwendige.“

szubjektivitás egyetemes és lényegi össz-szerkezete” (Husserl 2000. 70). Az emberi tudat genezise történetileg azt jelenti, hogy a konkrét (emberi) egóként objektiválódott transzcendentális szubjektivitás a transzcendentális filozófiák és filozófusok által bejárt történelmi-teleologikus úton halad egyetemes lényegi össz-szerkezetének, vagyis az észnek a megvalósulása felé.

A tudat genetikus története olyan fejlődés, amelyet egyszerre szemlélhethünk úgy, mint ami feltétele az igazság történetének, és úgy is, mint ami annak célja, sőt úgy is, mint ami az igazság történetének a célja. Ezért a fejlődés a tudatnak nem csupán lehetőséget biztosít arra, hogy a világot mint saját teljesítményét *értse meg*, de követelményként támasztja vele (tulajdonképpen magával) szemben, hogy saját teljesítményeként *hosszon létre* egy világot, vagy akként alakítsa át a világot, hogy saját teljesítményeként ismerhessen rá. Ennek a követelménynek nem felelhet meg akármilyen tudati tevékenység. Az *európai tudományok válsága* című írásában Husserl a következőképpen reflektál erre a kérdésre: mivel az ész megvalósulása a fejlődés telosza, a világmegérett és világgalkotó tevékenységből *szükségekppen* maradnak ki azok a szubjektumok, akiknek e két értelemben vett világkonstitúciós tevékenysége – appercepció rendszerük habitualitásának torzulása miatt – nem illeszthető be az ésszerűség teleológiai rendjébe. Ilyenek az elmeháborodottak, illetve a gyerekek, akik állítólag nem végeznek ilyen konstitúciós tevékenységet, csak – mások közvetítésével – a világ elsajátítására képesek. Őket, és velük együtt még az állatokat is, a normális-normatív világkonstitúciós tevékenységhez *mérten* sajátos világkonstitúció jellemzi, amelyet analógia alapján módosult intencionalitásként írhatunk le. Az analógia miatt mégiscsak bekapcsolódnak az észmegvalósulás teleologikus rendjébe, ám éppen mint nem normálisok, s ilyeneként mint (aktuálisan) nem normatív érvényű konstitúciót végzők. Ez az analogikus sor azonban nem áll meg az állatoknál, hanem „végül valamennyi élőlényt” magába foglalja, „amennyiben ezeknek van valamiféle »életük«” (Husserl 1998. 235). Ezért ha a normális-normatív világkonstitúció történetét vizsgáljuk, a „transzcendentális visszakerdezés” minden élőn végighalad, mert csak így tárulhat fel *a tudat genetikus története*.<sup>4</sup>

Vajon a módosult intencionalitások kontinuuus sorával, amely minden élőt magában foglal, nem ugyanahhoz jutottunk-e, amit fentebb elvetettünk, vagyis a biológiai evolucionizmushoz? A tudat geneziséét kutató transzcendentális vissza-

<sup>4</sup> Kérdéses persze, hogy milyen jellegű és milyen mértékű módosulást tűr el a világkonstitúcióban szerepet játszó normális intencionalitás anélkül, hogy kirekesztődne a világkonstitúcióból. Egyáltalában milyen szükségszerűség dönti el, hogy mely habitualitások illeszkednek és melyek nem az ésszerűség rendjébe? Az sem teljesen világos, hogy mi jelöli ki az ész teleologikus rendjét, vagy röviden a teleologikus rendet, és hogy honnan ered a célszerűség. És külön vizsgálatot igényel, hogy vajon az ész önmegevalósulása nem feltételezi-e a habitualitások folytonos változását, melyek közül adott pillanatban éppen azok nem illeszkednek, mert nem szabad illeszkedniük a fennálló rendbe, melyek a későbbiekben „előrevivőnek” bizonyulnak. Most egyelőre tekintünk el ezeknek a problémáknak a tárgyalásától, hogy majd később visszatérjünk hozzájuk.



kérdés módszertanával – amely arra utal, hogy nem az emberi tudat evolúcióját kutatjuk, hanem *a tudat* transzcendentális alapjaira kérdezzük vissza – sem lehet elleplezni azt a schopenhaueri belátást, hogy az anorganikus és organikus világot, illetve az organikus anyagot és az emberi tudatot kontinuos evolúció köti össze, és az evolúció során objektiválódott alakzatok egymás hatványozott analogonjai.

Nem kétséges, hogy Husserl a genetikus fenomenológiai elemzésekkel nagyon közel kerül, legalábbis témáját illetően, a biológiai evolucionizmus kérdésfeltevéséhez. A *Karteziánus elméletek*ekben például így összegzi a genetikus problémák szintjeit a transzcendentális fenomenológiában:

A pszichofizikai világrajövetel elvezet bennünket a tisztán biológiai-tes-ti egyedfejlődés, illetve filogenezis problémájához, melynek párhuzama a lélektani filogenezis. [...] Vajon mindebben nem a konstitúciós fenomenológia transzcendentális filozófiájának lényegi problémái bukkannak föl? [...] a genetikus problémák már a valóban fenomenológiai nevezhető munka legelső, legalapvetőbb szintjén is erőteljesen megjelennek. [...] Csakhogy ezzel a fentebb említett generatív problémákat, a születés és a halál, az animalitás nemzedéki összefüggéseinek kérdéskörét még nem is érintettük, hiszen magasabb dimenziókhoz tartozván az alsóbb szférák hatalmas kifejtő munkáját igénylik (Husserl 2000. 159).

Milyen értelemben nevezhetőek még ezek az elemzések *transzcendentál*-fenomenológiaiainak? Hiszen a transzcendentális elemzések lényege épp abban áll, hogy *amíg csak lehet*, biztosítsa a szellemi világ önállóságát és eredendőségét. A kérdés az, hogy meddig lehet ezt tenni. Komolyan vehetjük-e, hogy a transzcendentális visszakérdelésben feltárul és egymás alapjául szolgáló elő-, ő-, őseredeti rétegek hagymahéjként záródnak egymásra? Nem képez-e határt (végpontot és egyben korlátot is) a *transzcendentális* elemzésekben a biológiai-fiziológiai testiség? Avagy ez a testiség része lehet a transzcendentalitásnak, és értelmes transzcendentális testről beszélni? Megelőlegezve a későbbieket, tömören csak annyit válaszolunk erre a kérdésre, hogy igen, a testiség nemesak alávethető transzcendentális elemzésnek, de bizonyos értelemben maga jelenti a transzcendentalitást.

De mi az, hogy transzcendentális test(iség)? Egyáltalán, mi köze a fentebb elmondottakhoz a testnek? Akkor tudunk válaszolni erre a kérdésre, ha megvizsgáltuk, mi a szerepe a husserli fenomenológiában a testnek. Hangsúlyozom: a husserli fenomenológiában, hiszen tudjuk, hogy éppen a testiség (*Leiblichkeit*) Husserl által elvégzett fenomenológiai elemzéseinek kritikájából egy egészen más fenomenológia is kibontakozhat, és például Merleau-Ponty vagy Michel Henry elemzéseiben ki is bontakozott. Ha a testiség fenomenológiai leírására lennénk kíváncsiak, akkor természetesen ezeket az újabb fenomenológiai le-

írásokat is áttekintenénk. Most azonban nem azt vizsgáljuk, mennyiben találóak Husserl leírásai, hanem azt akarjuk megérteni, miként alakították át a sajáttestre (*Leib*) vonatkozó, általa is „újszerűnek” nevezett fenomenológiai belátások a fenomenológia néhány korai alapelvét, és miképpen vált a testiség a fenomenológia mint univerzális tudomány tervének teherhordójává. Egyértelműnek tűnik például, hogy a transzcendentális szubjektivitás fogalma (mondhatnánk fenomenális tartalma) épp a testiség elemzésével bővült ki jelentős mértékben (tett szert egyáltalán valamilyen tartalomra). Ezt az elmozdulást sematikusan úgy írhatnánk le, hogy a transzcendentális szubjektum a descartes-i típusú szubjektumtól a leibnizi monász felé mozdult el.

Amikor tehát a testre vonatkozó fenomenológiai elemzéseket a husserli kontextuson belül tartva vizsgáljuk, arra koncentrálunk, hogy azokat a feszültségeket tárjuk fel, melyek a fenomenológiának mint univerzális tudománynak a terve, illetve a sajáttest-elemzések belátásai között támadnak. Véleményem szerint ez utóbbiak alapjában teszik kérdéssé annak a lehetőségét, hogy a fenomenológia a filozófus radikális öneszmélésén túl a világra vonatkozó végső univerzális tudomány beteljesedett alakzata legyen. Husserl azonban szükségszerűnek tartotta, hogy ha magunkat mint transzcendentális interszjektív szubjektumokat átvilágítjuk, akkor az univerzális ész tudományához és az ennek produktumaként előálló igaz tudáshoz jutunk el. Ehhez valóban az kell, amit a mottónak választott idézet állít: vagyis hogy lehetséges legyen faktumot, ideális lehetőséget és szükségszerűséget tudományos viszonyba állítani. S mivel Husserl mindvégig ragaszkodott ehhez az elképzeléséhez, fenomenológiája az eredeti szándékkal szemben metafizikává alakult, vagy legalábbis ráépült egy leibnizi típusú metafizikára. Igaz, Husserl ezt a metafizikát megkülönböztette a történetileg elfajult metafizikától, és csak annyiban nevezte metafizikának, amennyiben a lét végső ismeretét, az első filozófiát annak szokás nevezni.

## II. TRANSCENDENTALITÁS ÉS TESTISÉG FESZÜLTSEGE

A címben jelzett probléma, vagyis transzcendentalitás és testiség viszonya feszültséggel terhes. Úgy érezzük, amit ez a két fogalom megnevez, az nem kapcsolódhat össze, épp ellenkezőleg, transzcendentalitás és test(iség) kizárják egymást. Ha egymás mellé rendeljük is őket, csak azzal a szándékkal tesszük, hogy kimutassuk, nem férnek meg egymással: transzcendentális test vagy testi transzcendentalitás – értelmetlen és értelmezhetetlen kifejezéseknek tűnnek. Ennek oka mindenekelőtt abban keresendő, hogy a transzcendentálishoz térképzetet társítunk, és létszféra<sup>1</sup>ként értelmezzük. Az így értett transzcendentalitás több konnotációt is hordoz. Az egyik alapján a *tudati immanencia* apodiktikusan evidens, tiszta szférájának tekintjük, mely éppen annak köszönheti apodikticitását és tisztaságát, hogy ki van belőle zárva minden transzcendens – a tudatimmanen-

ciához képest tudattranszcendens – tárgyiség, ami jelesül a test(iség)et jelenti. A másik szerint az *abszolút lét* szférájával azonosítjuk, mondván: minden, ami van, konstituáltként van, ekként pedig relatív az őt konstituálóhoz képest, mely így maga abszolút. Ennek következtében a test és a testi lét a maga egyediségében és esetlegességében nem tarthat a transzcendentális szférához. A harmadik konnotáció értelmében univerzális-örök szférának értelmezzük, melyben szükségszerűség uralkodik, egészen pontosan önkibontakozásának *teleologikus szükségszerűsége*, amivel nyilvánvaló ellentmondásban látszik lenni az önkibontakozásra képtelen, így mindenfajta teleológiát nélkülöző test(i lét).

Nem lehet nem észrevenni, hogy a fogalom Husserl használatában túlnyomórészt ezeket a konnotációkat hordozza. A transzcendentális fenomenológia a tiszta tudatban adott abszolút lét tudománya, mely magából a tudatból szükségszerűen bontakozik ki. Ez azonban nem jelenti azt, hogy Husserl csak ilyen értelemben gondolt volna a transzcendentalitásra, amikor transzcendentálisnak kereszteli el a fenomenológiát. Amiből persze nem következik, hogy rögtön valamilyen egészen más értelmet tulajdonított volna neki. Az hiszem, nem járunk messze az igazságtól, ha azt mondjuk, hogy a transzcendentális fordulat idején Husserl maga is ráhagyatkozott ezekre a konnotációkra, és rájuk bízta magát az elemzéseiben is. A fenomenológia mint univerzális transzcendentális tudomány eszméje pedig mindig is ezekre a jelentésmozzanatokra épül.

Amíg tehát a transzcendentalitás a tiszta tudat abszolút és szükségszerű szféráját jelenti, addig a test a maga anyagosságával, esetlegességével és meghaladhatatlan egyediségével nem tarthat ehhez a szférához. Ha viszont mégis, az csak azon az áron történhet, hogy a test(iség)re is kiterjesztjük a transzcendentalitás fentebbi jellemzőit, és tudattranszcendens helyett tudatimmanensnek, relatív helyett abszolútnak, végül esetleges helyett teleologikusan szükségszerűnek tekintjük. Ehhez azonban a transzcendentális nem maradhat egy létszféra neve. Talán éppen ennek felismerése motiválhatta Husserlt abban, hogy utolsó nagy művében, a *Válság*-könyvben a visszatekintő tisztázás és önértelmezés szándékával reflektáljon arra, mit is ért, vagy akar érteni transzcendentalitáson, és annak a fentiekől eltérő értelmét határozza meg:

Jómagam a *legtágabb értelemben* használom a „transzcendentális” kifejezést: azt a [...] motívumot értem rajta, amely Descartes közvetítésével valamennyi újkori filozófiát értelmessé teszi [...]. A megismerés képződményeinek végső forrására való visszakérdezés motívuma ez, a megismerő önmagára és megismerő életére való rácszémkedésé, amennyiben valamennyi, számára érvényes tudományos képződmény célszerűnek mutatkozik; mint saját tevékenysége eredményeit őrzi meg őket, amelyekkel szabadon rendelkezhet és rendelkezik. Radikális formában tisztán e végső forrásból feltörő, tehát végső alapokon nyugvó egyetemes filozófia motívumáról van szó (Husserl 1998. 130).

Még mielőtt elhamarkodottan úgy ítélnénk, hogy semmi sajátos nincs ebben a meghatározásban, semmi olyan, amiben eltérne a szó hagyományos értelmétől, hiszen egymás után bukkannak fel a tiszta forrás, az abszolút alap és a teleologikus szükségszerűség mozzanatai, vegyük észre, hogy egy momentum már itt is arról árulkodik, hogy a forrás, az alap és a szükségszerűség nem egy létszféra nevei vagy jellemzői. A *visszakérdés* motívuma a fentebb említett térképzetrel szemben a transzcendentális időbeli viszonylatban ragadja meg, tehát a Husserl által feltárt konstitúciós rétegek – például az én konstitúciójában – nem az én mélyére, mögé vagy fölé, hanem elé, korábbra vezetnek vissza.

Ennek következményeként a fenomenológiai leírás tárgya, a világértelem végső, abszolút, szükségszerű, transzcendentális forrása nem lehet a test nélküli tudat:

E forrás neve: *én magam*, egész valóságos és lehetséges megismerő lételemmel, végső soron *konkrét* [kiemelés tőlem – K. L.] lételemmel általában. Az egész transzcendentális problémakör középpontjában énem – az „*ego*” – áll, mégpedig ahhoz való viszonyában, amit eleinte magától értetődően mint vele azonosat tételeztünk: *saját lelkemhez*. De itt találjuk ennek az ének és önnön tudati lételemnek a *világhoz való viszonyát* is, amelyről tudomásom van, és melynek igazi létét saját ismereti képződményeimben ismerem meg (Husserl 1998. 130).

Mert az köztudott, hogy nem mindig ez az ego, a maga konkrét – a környező életvilágban való empirikus – életével volt a megismerés képződményeinek végső forrása. De nem is egy tetszőlegesen más ego. Kezdetben a fenomenológiai redukció egybeesett a descartes-i módszeres kétellyel; a redukció eredményeképpen egymástól élesen elválasztva egyfelől a fizikai természet, másfelől pedig a tiszta tudat állt elő. Mivel a tiszta tudat hordozója, a konkrétan létező én is része a világnak, ezért a redukcióban ő is zárójelbe került, így jött létre a világ nélküli, a maga benső világában önmagával beszélgető, minden személyességétől, sőt énségtől mentes tiszta tudat.<sup>5</sup> Ám bármennyire is igyekezett Husserl ennek a tudati szférának a tisztaságát garantálni, ügyelve arra, nehogy Descartes-hoz hasonlóan beemelje a tudatimmanenciába a világ egy darabkáját, észre kellett vennie, hogy a tiszta tudatban a világ fenomenális jelenlétén túl jelen van valami kiiktathatatlan világi elem, mely az észleléshez és minden észlelethez determináns módon hozzákapszódik. Ez a valami nem hagyja meg a tudatot a maga bensőségességében. A megismerés termékei ugyanis nem az öntudat a priori formáiban találják meg eredeti forrásukat, hanem a konkrét megismerő

<sup>5</sup> Ezzel kapcsolatban lásd Derrida kritikáját a *La voix et le phénomène* című művében (Derrida 1967), valamint a jelen írás szerzőjének *A fenomenológia alapvetésének derridai dekonstrukciója* című tanulmányát (Kicsák 2000).

tevékenységet végző énben. Az észlelés alapját jelentő érzéki észleletek utalnak a keletkezésük módjára is. Olyan érzetekről van szó, melyek értelméhez szükségképpen hozzátartozik a mozgás, ami viszont testet feltételez. Vagyis a test bejelentkezik a fenomenológiai szférába, mint annak kiiktathatatlan része, felülírva, kitágítva a transzcendentális addigi értelmét is.

Ettől a fentebbi idézetben meghatározott transzcendentalitástól nemcsak hogy nem idegen a testiség, de – ha Husserlhez hasonlóan a transzcendentalitást nem mint létszférát, hanem mint időben kibontakozó funkciót, vagyis mint azt a konstituáló működést értjük meg, melynek eredménye a konstituált transzcendens világ – azt is mondhatjuk, hogy a testiség alkotja magát a transzcendentálist. Ha a megismerés termékeinek végső forrása a konkrét ego a maga konkrét megismerő életével, akkor végső soron a testisége, mely az ego konkrétságának a lehetőség-feltétele, teszi lehetővé számára azt is, hogy egyáltalán ismerettel rendelkezék a világról.

### III. A TEST BEEMELÉSE A TRANSCENDENTÁLIS FENOMENOLÓGIÁBA – A KINESZTÉTIKUS TUDAT

Tudjuk, milyen lényegi szerepet játszik az észlelés a fenomenológiában. Husserl a *Fenomenológiai pszichológia* (1924) szisztematikus részében (39. § *Az észlelés elemzése magára az észlelőre való tekintettel*) egyértelműen beszél arról, hogy az észlelésnek a megszokott, és nemcsak ebben a műben, hanem általában a korábbi írásokban szereplő leírása szörnyen egyoldalúan történt (vö. Husserl 1968. 196–200: „Roppant mértékű [*ungeheure*], legszívesebben azt mondanám, rémes [*ungeheuerliche*] egyoldalúság jellemzi”). Miben állt az egyoldalúság? Abban, hogy az észlelést nem mint magát a folyamatot (*Wahrnehmen*), hanem csupán mint ennek a folyamatnak az eredményét (*Wahrnehmung*) vette tekintetbe. Elhanyagolta a noétikus oldalt a noématikus tartalom javára, vagyis az objektumra, a tárgyi megjelenések lefutására koncentrált. Az észlelés ebben az egyoldalú értelmezésben annyit jelentett, mint a megjelenésben megjelenőt, a perceptív módon véltet birtokolni. Az észlelés eredménye azonban „nem az égből pottyann le”, csakis az észlelés folyamatában teszünk szert rá. Az észlelés folyamatának pedig megvannak a sajátosságai, melyeket *szükségszerű funkciójukban* lehet tanulmányozni, azaz transzcendentális elemzésnek vethetünk alá. Husserl többször is hangsúlyozza, hogy az elemzések megmaradnak a szubjektív bensőségben, a belső intencionális szférában. Azért van erre szükség, mert most véterik fel az egyediesetleges test az apodiktikus szükségyszerűség közegébe.

Lássuk, milyen lépéseken keresztül. A dolgok jelenségeit vizuálisan a látóban, taktilisan a tapintóban, és így tovább, észleljük; csak ezáltal teszünk szert erre vagy arra a jelenségre. Ebben a faktumban szükségyszerűség rejlik, mely lényeginek mutatkozik. Ugyanis minden észlelet észlelőre utal, és nem valami-

lyen általános, hanem mindig konkrét-egyedi észlelőre: a vizuális észleletek a látóra, a taktilisak a tapintóra. Ez az utalás nem egyszerűen kísérője az észleleteknek, hanem az észlelés lényegéhez tartozik. Az észlelés intencionalitását tehát, amely a dolgot észleltként adja, nem lehet anélkül tanulmányozni, hogy ne vizsgálnánk azt az intencionalitást is, amely a sajáttest észlelőfunkcióinak felel meg. Az észlelésben ezért kétfajta intencionalitás, a tudati és a testi intencionalitás épül egymásra. Így megkerülhetlenné válik, hogy tematikus vizsgálat alá vegyük az észlelőszervek egységét, vagyis a testünket.

A test fenomenológiai megközelítését megnehezíti, hogy a szemlélet eltárgyasít, a sajáttestet viszont nem tárgyként kell megragadnunk, hanem olyan észlelőfunkciók intencionális egységeként, melyek tárgyakat adnak, vagy adják a tárgyakat. Lehetséges ez? Igen, mert a sajáttest rendelkezik azzal a specifikummal, amellyel semmilyen más test nem: úgy mutatkozik meg nekem, tehát oly módon válik intencionális aktus tárgyává, hogy közben nem tárgyasul. A saját testünk számunkra elsődlegesen és közvetlenül nem mint szemléleti tárgy van adva, hiszen ekkor maga is jelenség lenne, hanem mint funkció (intencionális működés), ebben az esetben észlelési funkció. Ez a legkritikusabb pontja az elemzésnek, és ezt Husserl is tudja, ezért beszél a probléma „felelősségteljes tudományos kibontásáról”. Mert nem elegendő annyit megállapítani, „pszichológia előtti megfontolásokból” kiindulva, hogy a testünk egyszerre dologként és egyszerre funkcióként, specifikus testként is adott; fenomenológiailag is hozzáférhetővé kell tenni, és le kell írni a sajáttest tapasztalatát. Husserl szerint ez is fenomenológiai redukcióban történik meg, analóg módon a külső észleléshez, hiszen a test is térbeli dologként adódik. Azonban az analógia itt véget is ér, mert a test több tisztán térbeli dolognál. A saját test észlelése alapvetően újszerű észleleti értelmet és vele egy újszerű észleléstípust jelent.

Az új itt az, hogy ez az észlelés, jóllehet térbeli-dologi észlelés formája van, egészen újszerűt nyújt, és vele egy újszerű észleléstípust is, ahogyan a korlatívan transzcendensen vélt és tételezett, ami itt a test, maga is alapvetően újszerű (Husserl 1968. 198).

Az itt az újszerű, hogy nem csupán a specifikus testi intencionalitás tárul fel, hanem ennek egybeszővődése a tudati intencionalitással. A tudati és testi intencionalitás kölcsönösen egymásra vonatkozó intencionalitás, és amit eddig, egyoldalúan, tisztán tudati intencionalitásnak tekintettünk, az nem csupán indexként, hanem lényegi determinánsként hordozza magában a specifikus testi intencionalitást, a kinesztézist. A kölcsönös egymásra utalás és egymásra utaltság okán önmagában a tudati, illetve önmagában a testi intencionalitás csak absztrakcióval válik szét egymástól. *Az észlelés maga kinesztétikus intencionalitás. A test transzcendentális meghatározása azonos a kinesztétikus tudattal.*

Ne gondoljuk azonban, hogy ezzel a test „felszívódik” a tudatban, mert a testet magát kétoldalúság, egyfelől szubjektív bensőség, másfelől objektív külsőség jellemzi. Egyidejűleg, hiszen a bensőtestiség nem más, mint szubjektív működési módok potenciálisan adott habituális rendszere, ami külsőleg szervekben és objektív térbeli mozgásokban aktualizálódik. Vagyis ami a kinesztétikus tudat testi meghatározottságát jelenti, az a test objektív térbeli mozgásának függvénye.

A test azonban nemcsak az észlelő tevékenységben játszik döntő szerepet a szubjektív bensőtestiséggel és az objektív térbeli mozgással együtt, hanem mint a világba beavatkozó, a természeti folyamatokkal objektívan változó is. Vagyis nemcsak szemlélő, személetet nyújtó, hanem gyakorlati-cselekvő szerv is, melyet az én *hégemonikon*ként mozgat.

Husserl azzal zárja ezt az okfejtést, hogy a testi intencionalitás elemzésének eredményei olyan új felismerésekhez vezetnek, amelyek a térbeli világ adottságának fenomenológiája során nélkülözhetetlenek. A világnak mint térbeli világnak a lehetősége ugyanis a kinesztétikus tudat alapját alkotó és orientációs rendszerünk nullpontjaként funkcionáló testben rejlik.<sup>6</sup>

### 1. Testek

Ha most arra a kérdésre keressük a választ, hogy miként kell érteni és hogyan lehet megragadni ezt a tranzsczendentális-fenomenális testet, akkor fel kell vázolnunk azt a teljes konstellációt, amelybe az *Fiszmék* II. részében a test – a természet és a szellem között – a lélekkel együtt elrendeződik.<sup>7</sup> Előzetesen má-

<sup>6</sup> Az a gondolat, hogy itt nem csupán az eddigi leírások egyoldalúságának kiegyensúlyozásáról, kiegészítéséről van szó, vagyis hogy az újszerű belátások a fenomenológia alapjait érintik, ott lappang ugyan az elemzésben, mégsem tör át. Nem válik kérdésessé például, hogy a sajáttest tapasztalata valóban analóg-e a térbeli dolgok tapasztalatával. Husserl nemcsak evidensnek veszi ezt, de a test tapasztalatát olyan észlelésként határozza meg, amely szükségképpen intencionális viszonyban valósul meg. Kiemeli, hogy ez az intencionalitás nem azonos a szemléletben megjelenített tárgyával. Ám arra, hogy mi áll ebben a viszonyban a szubjektív póluson, ha egyszer a test a tárgyi pólus, nem kérdez rá. Nem kérdez rá arra sem, hogy miben állnak az újszerű észleleti értelemek, és mi jellemzi azt a bizonyos újszerű észleléstípust. Nem másról van itt szó, mint arról, hogy a test, amely minden észlelés feltétele, most az észlelésünk tárgyává válik. De nem egyszerű tárgyként jelenik meg, hanem funkcionalitásában. Vagyis az éppen észlelő testet kell(ene) észlelnünk. A sajáttest észlelése esetében azonban megbillen az az elrendezett viszony, amely a tudatként felfogott tranzsczendentális szubjektum (a működő, konstituáló szubjektum) és az őt szemlélő fenomenológiai szubjektum (az elemzést végző fenomenológus) között épül ki. A működő tudatszubjektum „fölé” tudok emelkedni, a testből viszont nem lehet kiemelkednem, mert a test észlelése testileg meghatározott észlelést jelent. Vagyis észlelőként ez a test soha nem észlelhető. Nem mellesleg ezek Merleau-Ponty és Michel Henry legfőbb ellenvetései Husserlrel szemben.

<sup>7</sup> „Also haben wir z w e i P o l e: physische Natur und Geist und dazwischen Leib und Seele.” Husserl 1984. 116.

ris szögezzük le: a fenomenális test a konkrét hús-vér fizikai testem, azonban nem fizikaként szemlélve, azaz nem mint tőlem független, külső észlelésem tárgyaként adott dolog, hanem mint olyan test, *melyben én működöm*. Ez a konkrét hús-vér test különböző létrétegek megbonthatatlan, komplex egységeként konstituálódik, melyek – csakis a leírás kedvéért – analízissel, redukcióval és absztrakcióval különválaszthatók egymástól.

Az első megkülönböztetés a beállítottság-váltás alapján megy végbe. A természeti-természettudományos beállítottságban a test mint téridőbeli objektum lesz szemlélet tárgyává: ezt nevezi Husserl *fizikai test*nek. Ekként adott számunkra, külső észleletként, minden test a maga fizikai tulajdonságainak összességéeként, és ekként adott mások számára, illetve a tudományos vizsgálódás számára a saját testem is. A magam számára a saját testem, még ha tudományos vizsgálat alá vetem is, akkor sem lehet pusztán fizikai objektum. Tárgyként viszont adódhat, mégpedig a természetes beállítottságban, mint „a személyes környező világ jelensége és tagja”. Ez az *anyagi test*. A testem mindkét esetben konstituált objektumként válik hozzáférhetővé.

A fenomenológiai beállítottságban a test nem mint konstituált, hanem mint konstituáló egység jön tekintetbe. Ez a fenomenológiai test kétoldalú, amitől a személyes, konkrét én, éppen testi mivolta miatt, besugárzási és kisugárzási centrummá (*Ein- und Ausstrahlungszentrum*) lesz. A test *egyfelől* érzékelő mező, s mint ilyen affektusoknak passzívan kitett és ezek felfogására képes érzőtest (*aesthesiologischer Leib*), *másfelől* pedig szabad mozgások végzésére, aktivitásra képes, akaró-test (*Willensleib*). Az előbbi értelemben a fizikai, helyesebben a materiális testtől függ, vagyis a természetbe ágyazódik, az utóbbi értelemben pedig a szabadon tevékeny szellemnek van alárendelve.

Ennek következtében a test közvetítő médium a külső és belső világ között, a természet és a szellem között, amit a fent említett kétarcúságának köszönhet. Egyébként a természet és a szellem világa (a test és a lélek) közötti viszonyt Husserl feltételeességi viszonyként határozza meg. Ezáltal igyekszik a fenomenológia test-lélek felfogását kiemelni az addigi hagyományos elképzelések köréből. Ez a feltételeességi viszony azt jelenti, hogy a két szféra kapcsolatára nem alkalmazható sem a hagyományos paralelizmus, miszerint a szellemi folyamatokat velük egy időben lezajló és nekik megfelelő testi folyamatok kísérik, sem pedig a természeti világra jellemző kauzalitás. Így a szellemi világ és a természeti világ között a kauzális kapcsolat hiányában sem reduktív, sem deduktív viszony nem lehetséges.

(Ezzel Husserl permanens paradoxiaiban tartja a testproblematikát. A célja ugyanis a pszichofizikai paralelizmus felszámolása, aminek alapját a descartes-i dualizmus képezi. Ez a meghaladás azonban nem vezethet el egy spinozai típusú monizmushoz, hiszen akkor a szellemtudományok, önálló tárgy hiányában, lehetetlenek lennének. Természet és szellem, anyag és szellem, test és szellem dualitását fenntartva kell tehát meghaladni a redukcionista-deduktív tendenciákra hajlamosító descartes-i dualizmust.)



## 2. Test és igazság

Fentebb láttuk, hogy a kinesztétikus észlelés fenomenális teste a térbeli orientációs rendszerünk nullpontja, abszolút itt-és-mostként redukálhatatlan egyediség. Miután a test az észlelés lehetőségfeltételének bizonyul, a transzcendentális szubjektivitás meghaladhatatlanul egyedivé válik. Az univerzális tudomány alapjait ezeknek az egyedi észleleteknek kell alkotniuk úgy, hogy az egyediségből utat találunk az univerzalitáshoz, a helyhez és időhöz kötöttségből az időtlen igazsághoz. Az egyediség csak akkor nem zárja ki a szükségszerűséget, ha maga az esetlegesség válik szükségszerűséggé, ha a test alá van vetve annak a transzcendentális teleológiának, mely a világ végső igazságához vezet. Most rekonstruáljuk azt a két lépést, melynek során a test ennek a transzcendentális teleológiának rendelődik alá.

A) *Normalitás – anomáltság*. Kiindulásként térjünk vissza egy korábbi idézethez. Emlékszünk, Husserl azt írja az *Eszmék* II. részében, hogy ha az egyének a jelenségekre vonatkozóan egybehangzó ítéletet alkotnak, az véletlenszerű tény, ha ellenben a tapasztalatot ésszerűen meg gondolják, és úgy jutnak megegyező ítéletre, az szükségszerű:

Ha a közösségi szubjektumok az individuális vagy interindividuális módon megjelenő tárgyak esetében a jelenségről egybehangzó ítéletet alkotnak, az csupán véletlenszerű faktum, ha ellenben a tapasztalatot feldolgozó ésszerű gondolkodás során az objektív (fizikai) meghatározottságok tekintetében jutnak megegyezésre, az szükségszerű (Husserl 1984. 38).

Majd úgy folytatja, hogy az eltérések és az egybeesések – vagyis a jelenségek észlelése során jelentkező szükségszerű eltérések és véletlenszerű egybeesések, illetve a tapasztalat ésszerű gondolati feldolgozása során jelentkező véletlenszerű eltérések és szükségszerű egybeesések – maguk nem véletlenszerűek, hanem nagyon is szabályozottak. Megvan a pontosan feltárható okuk, mégpedig az egyének testi és lelki organizációjában: az egybeesések oka ennek az organizációnak az azonossága, az eltérések okai pedig az organizációban jelentkező különbségek.

Ehhez járul még továbbá az is, hogy ennek az objektív teoretikus természetismeretnek a segítségével a szubjektumoknak a jelenségekben adódó *minden valóságos vagy lehetséges eltérését és megegyezését* (a dolgok közvetlenül tapasztalt állapotát) pontosan a szubjektumok objektíve kutatható *testi és lelki organizációjára való tekintettel lehet megmagyarázni*, tehát maguk is *szükségszerűnek* bizonyulnak (Husserl 1984. 38, kiemelés tőlem – K. L.).<sup>8</sup>

<sup>8</sup> „Dazu gehört des weiteren, dass mittels dieser objektiven theoretischen Naturerkenntnisse alle wirklichen und möglichen Differenzen und Übereinstimmungen der Subjekte in ihren

Létezik tehát egy olyan testi-lelki organizációs állapot, mely szükségszerűen vezet el az egyedileg eltérő észleletektől az ésszerű gondolatok megegyezéséig (*Übereinstimmung*). Ezt nevezi Husserl *normalitás*nak. A tapasztalatot megragadó ésszerű gondolatok ugyanakkor nem csupán egyedi tudatok korrelátumai, de „korrelátumai” az objektív világnak is. Vagyis a közös gondolatok nemcsak egymással egyeznek meg, de egybeesnek a világ igazságával is, helyesebben konstituálják azt. Röviden azt mondhatjuk: igaznak a normális testiségben optimálisan adottat nevezzük.

A „normális” testiséghez hozzátartozik a megjelenítés lehetőségeinek jól vagy rosszul működő rendszere, és ehhez továbbá hozzátartozik egy optimális adottságmód eszméje, vagy adottságmódok lehetséges optimális rendszere, mely egy adott fokon az igazságot képviseli (Husserl 1973. 121).<sup>9</sup>

A fentieket összegezve, azt találjuk, hogy a világ nem más, mint egy testre vonatkoztatott lehetséges téridőbeli tapasztalatok univerzuma. A test egyedisége okán a világról adódó szemléletek redukálhatatlanul különbözőek. A testiséggel jelentkező, meghaladhatatlan egyediség és a vele járó egyedi világszemlélet ellenére a kommunikatív konstituálódó életvilágban egy általános-közös világ konstituódik. Hogyan lehetséges ez? Husserl szerint úgy, hogy *a biofizikai normalitásnak és anomaliásnak megfelel egy konstitutív normalitás és anomaliás*. Minden normális embert elsődlegesen normális testiség jellemez, normális érzékszervekkel és normális érzékelőmezővel, vagyis normális pszichikummal. Mint minden konstitúciós folyamatnál, természetesen itt is kétpólusú korrelatív viszonyra bukkanunk: A *normálisan működő szemlélő* (vagy a szemlélő normális működése) korrelátumként *az igaz módon szemlélt világot* (vagy a világ igaz szemléletét) teremti meg. Az egyik póluson tehát a normálisan működő test vagy a testi működés normalitása áll, melynek kapcsán Husserl eredendő normalitásról beszél, és ezt a testi viselkedés egy bizonyos tipikus állandójának tartja. A tipikusság és az állandóság a percepciós és appercepciós képességek habitussá váló aktualizációjában nyilvánul meg. Ha ez az állandósult viselkedés nem juthat érvényre, a világ elbizonytalanodik, és amíg a bizonyosság helyre nem áll, látszatként marad érvényben. A helyreállást a korrelatív viszony másik pólusa, az optimális adottságmód vagy adottságmódok lehetséges optimális rendszere biztosítja.

---

Phänomenen (den unmittelbar erfahrenen dinglichen Beständen) unter den Rücksichtnahme auf die ebenfalls objektiv erforschbare leibliche und seelische Organisation derselben zu erklären, also selbst als notwendig erkennbar seien.”

<sup>9</sup> „Zur »normalen« Leiblichkeit gehört ein System von Möglichkeiten der Darstellungen, schlechter und guter, und diesem wieder gehört zu die Idee einer optimalen Gegebenheitsweise oder eines möglichen optimalen System von Gegebenheitsweisen, das auf einer gewissen Stufe die Wahrheit vertritt.”

Rövid kitérőt kell most tennünk, mert súlyos problémák merülnek fel az elmondottakkal kapcsolatban. Erősen kérdéses mindenekelőtt az, hogy az optimális adottságmód vagy az adottságmódok optimális rendszere az észlelő biofizikai organizációjának és a tárgy fizikai megjelenésének az összhangjában áll-e. Mert amit igazságnak, a világeszlelés és a világról alkotott ítéletek igazságának, vagyis az adottságmódok optimális rendszerének nevezünk, *szokásszerűen aktualizálódó képességek* működéséhez van kötve. A habitualitás, a típusosság olyan fogalmak a husserli fenomenológiában, amelyek többet fednek el annál, mint amennyit feltárnak, miközben lényegi pontokon nagyon komoly teherhordó elemei a fenomenológiának. A legtöbb helyen leíró fogalmakként működnek, és az introspekción alapuló, szükségképpen egyedi tudatelemzések általános érvényűségének a lehetőségfeltételét jelentik. Ezen a ponton azonban, vagyis a normalitásról szóló leírásokban lényegileg nem maradhatnak meg csupán leíró fogalmaknak, ahogy egyébként, mint mindjárt látni fogjuk, maga a normalitás sem. Egyáltalán nem mindegy, hogy a „képes vagyok” éppen milyen mozgási, észlelő, megismerő és gondolkodási folyamatokban aktualizálódik, de nem abban az értelemben, hogy erről a tárgy megragadásának sikeres vagy sikertelen esetei döntenének. Az aktualizációnak meghatározott módon *kell* megtörténnie, mert a világ az ettől eltérő esetekben a látszat érvényességét hordozza. Vagyis a habitualitás nem bármilyen intencionális ráirányulások szokásszerű rendszere, hanem csak bizonyos kiválasztottaké, melyek követendő normákká válnak, és az egyéneknél szokások kialakításával állandósulnak. A típusosság ezekben a történetileg változó, kondicionálással interiorizált normák formájában gyakorolt intencionális viszonyulásmódokban figyelhető és állapítható meg.

Husserl szerint ezt a kiválasztást az említett korreláció világi pólusa, az optimális adottságmódok rendszere vezérli. A normalitás azonban kommunikatív közösségbe tömörült személyek sokasága között áll fenn, akik a tapasztalataikat és a tapasztalati kijelentéseiket illető szabályok felől átlagosan megegyeznek,<sup>10</sup> szemben ugyanezen közösség más tagjaival, akik környező világukat, életvilágukat egyedi irányultságú leírásokban ragadják meg, és a „mindenkori motiváció keretei között empirikusan nem megvalósítható módon tapasztalják” (Husserl 1984. 38). Hogy mik alkotják a motiváció fennálló kereteit, arról egy közösség normálisnak tekintett tagjai döntenek, miközben ez normalitásuknak a feltétele is egyben. Normális az, amiben a közösség normális tagjai mint az egyedi szubjektív észleletek átlagában megegyeznek, abban az értelemben, hogy ítéleteiknek van közös tartalma, ami egyben egy intézményalapító megegyezés – konvenciókötés – alapját is képezi.<sup>11</sup> Ezt az átlagot a normálisok határozzák

<sup>10</sup> Az *Übereinstimmung* másik értelme vezérli az értelmezésünket.

<sup>11</sup> Miről van szó? Eleveintsük fel a tapasztalati ítéletek egybeesésére vonatkozó érvelést, és képzeljük el, hogy két normális észlelőről van szó. A különböző szubjektumok számára a dolgok természetszerűleg különböző módon jelennek meg abból a tényből eredően, hogy maguk téridőbeli (tehát fizikai-testi) mivoltukban elkülönülnek egymástól, nem lehetnek

meg, de a norma a megegyezésben teremődik. A normalitás nem valamilyen természettől adott szükségszerűség tehát, hanem tulajdonképpen megegyezésen alapuló konvencionálisnak, azaz normativitásnak bizonyul. És elsődlegesen az anomálitás sem a testi-lelki organizációban jelentkező különbség, hanem ettől a konvencionális normativitástól való eltérés. Mert hogyan is lehetne biofizikai különbségekre észleleti és észleletítéleti különbségeket visszavezetni, tekintve azt a végtelen eltérést, ami egy testi organizmusban fiziológiai szinten kimutatható? Hogy érzékeljük, milyen súlyos problémák rejlenek itt, gondoljuk meg a következőket: a világkonstitúció és még azt megelőzően egyáltalán az értelemképződés során jelentkezhetnek olyan anomáliák, melyeknek nem felel meg semmilyen szomatikus vagy pszichikai anomália. Ha azt el is fogadjuk, bár korántsem evidens, hogy pszichofizikai anomáliák szükségképpen értelem-anomáliákhoz vezetnek, az viszont egyáltalán nem biztos, hogy a pszichofizikai normalitás kizárja az értelem-anomáliákat.

Az a gondolat, hogy ez a megegyezés nem egybeesést, hanem konvenciót alkotó szerződést jelent, nem bukkan fel Husserlnél. Az pedig, hogy a később tévedéseknek bizonyuló valamikori igazságok – melyekre éppúgy érvényes, hogy a magukat normálisnak tartó személyek közösségének interszubjektíve konstituálódó korrelatívumai – hogyan és miért változhatnak meg, nem válik kérdéssé. Ahogyan az sem, hogy melyik változik előbb, vagy hogy milyen mértékű változás vezethet el az igazság megváltozásához, és hogy mennyiben múlhat az igazság megváltozása a habitusok véletlenszerű megváltozásán.

Mivel Husserl a megalapozás eszméje vezérli, ennek a normalitásnak a testi-lelki organizáció természet adta működésében találja meg az alapját. *A természet normát teremt.* Husserl szerint a természetkutatás során konstituált fizikai

---

ugyanabban az időpillanatban ugyanazon a helyen. Tudjuk, ha ennek ellenére egybehangzó ítéletet alkotnak, a közvetlenül tapasztalt érzéki-szemléleti jelenségről, az pusztán véletlenszerű egybeesés lehet. Szigorú értelemben azonban nem csupán véletlenszerű, de lehetetlen is, hogy az ítéleteik megegyezzenek, ezért az egybehangzó ítéletek közül valamelyiknek hamisnak kell lennie ahhoz, hogy megegyezhessen egy másikkal. Az egyik szubjektumnak tévednie kell, tévesen kell érzékelnie a jelenséget vagy tévesen kell megítélnie az érzékelt jelenséget. És itt komoly nehézségre bukkanunk, hiszen vagy azt kell feltételeznünk, hogy a jelenség mint megjelenés önmagában lehet téves, tehát egy fenomen adódhat nem igaz módon, vagy azt, hogy az ítélet függetleníti magát az érzéki szemlélettől. Az előbbinek a lehetőségét önmagában ki kell zárunk, az adódás tudniillik olyan végső és meghaladhatatlan feltétele a fenomennek, amelyről nem mondhatjuk, hogy lehet igaz vagy hamis. Az adódás van, és ennek függvényeként van a fenomen. Vagyis a tévedés azért lehetséges, mert *az ítélet függetlenedik az érzéki szemlélettől.* Ha viszont függetlenné válhat, akkor az igazsága sem abban áll, hogy egybeesik vele. A jelenségre vonatkozó igaz ítélet nem a jelenséggel egyezik meg, hanem a megjelenés optimális rendszerébe illeszkedik.

Amit a jelenségre vonatkozó azonos ítéletnek nevezünk, azt a szavak adekvát használatra teszi lehetővé. A szavak, a nyelv bázisa az, amit Nietzsche a nem azonos azonossá tételének nevez, a különbség eltörlésért eredményező nivelláció. Ekkor azonban a testi-lelki organizációban jelentkező különbség vagy megegyezés nem lehet az ítéletben jelentkező különbség és egybeesés szükségszerű oka. Márpedig Husserl erre alapozza a normalitásról szóló okfejtését, amin pedig azután az igazság teleologikus történetének elképzelése nyugszik.

dologi objektivitás ugyanaz, mint ami egyfelől az egyedi szubjektumok számára személetesileg konstituálódik (mintegy magától, passzív-spontán módon), majd pedig leíró kijelentéseikben kifejeződik, másfelől pedig mint amelyik annak a közösségnek a korrelátumaként konstituálódik, amely a magukat normálisnak értékelő személyekből jön létre. Természetesen a szóban forgó szubjektumok ugyanazok. Az számít normálisnak, akiben szubjektíve személetesileg ugyanaz a dologi objektivitás konstituálódik meg, amelyet a magukat normálisnak tartó személyekből álló közösség interszjektíve konstituál. Ez nem véletlen, ez teleológia.

Mindent felülír a szükségszerűség. Ám ha mindez a testi-lelki organizációban van megalapozva, akkor az igazság megváltozása korrelátumként ennek az organizációnak a megváltozását feltételezi, és egyben eredményezi is. Husserl nem habozik ezt a megoldást választani: a magasabban fejlett testek jutnak el a magasabb rendű igazságokhoz, és a magasabb rendű igazság tudatosodása magasabb rendű testeket, azaz magasabb szintű testi szerveződéseket termel ki.

B) *A magasabb rendű testek organikus teleológiája.* A testhez kötöttség folytán tulajdonképpen minden *solus* világ anomáliát mutat, de megvan a lehetősége annak, hogy egyetértésre jussunk, kialakítsuk a világ észlelésének valamilyen egységességét. Mert bár a világkonstitúció egyéni, szubjektív rendszer terméke, de egy interszjektív rendszer által is meghatározott. A világhoz az is hozzátartozik, hogy nemcsak az én eredendő konstitutív rendszeremben áll elő ilyenként vagy olyanként, hanem mindenki más számára más és más módon is megjelenik, azonban ezek a különböző megjelenési lehetőségek egységgé, a valóságos és lehetséges megjelenések egységévé állnak össze. Létezik tehát egy interszjektíve lehetséges tapasztalati rendszer, melynek az egyik póluson a normális testiséggel működő egyének, a másik póluson a világról közösen megszerezhető pontos meghatározás eszméje felel meg. Mert az egyes szemlélet a priori mindenkor nyitva hagyja annak lehetőségét, hogy új szubjektumok lépjenek be abba az összefüggésbe, és a dologról még újabbat is megtapasztalhassanak.

Soha nem mondhatom, hogy egy tapasztaló szubjektivitás soha nem jelentheti ki bizonyossággal, hogy valaki ne tudna több meghatározástartalmat, ne tudná esetlegesen a világ addig rejtve maradt oldalát megtapasztalni, mégpedig olyanok, akikhez „magasabb” testiség tartozik (Husserl 1973. 122).

A világ lehetséges változatainak tehát a testiség lehetséges változatai felelnek meg: a homályos, látszatszerű, nem igaz érvényességét hordozó világnak egy tökéletlen testiség, az igaz világ szemléletének egy magasabb testiség. Ez a szubjektum szintjén a sajáttest vagy egyes testi funkciók megváltozását jelenti, interszjektíve azonban ezek a változások már az alacsonyabb és magasabb szintű testi szerveződések világába: az állatvilág fajaihoz vezetnek. És ezzel el-

jutottunk az organikus teleológiához, az igazság feltárására egyre alkalmasabb testek kifejlődésének eszméjéhez, mely tulajdonképpen megismérlése a biológiai evolucionizmus gondolatának, hiszen ezeknek a testi individuumoknak, fizikai organizmusoknak a fejlődését a létfenntartáshoz és az egyéni fejlődésükhöz szükséges kedvező feltételek segítik.

Ami alapvetően megkülönbözteti azonban a transzcendentális teleológiát a biológiai evolucionizmustól, az a célszerűség, s ennek egyértelmű meghatározása: a fejlődés a világról létrehozandó igaz szemlélet céljából zajlik. Így a fejlődés, bármelyik szinten szemléljük is, magának a test-lélek-szellem egységeként létező monásznak a fejlődése. Helyesebben egy monászrendszeré, melynek rendszerszerűségét az egy és ugyanazon célból zajló fejlődések összefüggésgysége alkotja, és amely „az egy és ugyanazon világ különböző lehetséges szemléleteit fokozatosan és fejlődésszerűen valósítja meg, felemelkedik az emberhez és végül a magában mind magasabb szintű intelligenciát és tudományt kibontakoztató emberhez” (Husserl 1973. 128). A fejlődés magának a monásznak a létalapja, mert egy monász nem létezhet másként, csak amennyiben kifejlődik, és egy világszemlélet csak mint a monász fejlődésének produktuma jöhet létre. Amit igaz világnak nevezünk, az a világról való szemlélet módján valósul meg, egyszerre a megismerés és a létrehozás értelmében. Ezt a kibontakozást pedig úgy kell elgondolnunk, mint a tudat önkifejlődését.

## IRODALOM

- Derrida, Jacques 1967. *La voix et le phénomène*. Paris, PUF.
- Husserl, Edmund 1968. *Phänomenologische Psychologie*. (Husserliana IX.) Szerk. Walter Biemel. Den Haag, Martinus Nijhoff.
- Husserl, Edmund 1972. A fenomenológia eszméje. In *Edmund Husserl válogatott tanulmányai*. Ford. Baránszky Jób László. Budapest, Gondolat. 27–110.
- Husserl, Edmund 1973. *Zur Phänomenologie der Intersubjektivität. Zweiter Teil: 1921–1928*. (Husserliana XIV.) Szerk. Iso Kern. Den Haag, Martinus Nijhoff.
- Husserl, Edmund 1984. *Die Konstitution der geistigen Welt*. (Text nach Husserliana IV.) Szerk. Manfred Sommer. Hamburg, Felix Meiner Verlag.
- Husserl, Edmund 1998. *Az európai tudományok válsága I*. Ford. Berényi Gábor – Mezei Balázs. Budapest, Atlantisz.
- Husserl, Edmund 2000. *Kartezianus elmékedések*. Ford. Mezei Balázs. Budapest, Atlantisz.
- Kicsák Lóránt 2000. A fenomenológia alapvetésének derridai dekonstrukciója. *Gond.* 25–26. 115–154.
- Nietzsche, Friedrich 1997. *A vidám tudomány*. Ford. Romhányi Török Gábor. Budapest, Holnap.

## „Szubjektív tárgyiség”

### Husserl fenomenológiai elmélete a test megalapozó szerepéről

#### I. BEVEZETÉS: A KONKRÉT SZUBJEKTIVTÁS

A husserli fenomenológia értelmezéseiben nem számít újdonságnak annak hangsúlyozása, hogy Husserl az 1920-as évektől a transzcendentális tudat elemzését elválaszthatatlannak ítéli a világ mint tapasztalati horizont elemzésétől. Arra azonban már nem minden értelmezésben derül fény, hogy ez a fokozatosan elmélyülő meggyőződés milyen változtatásokat implikál magában a transzcendentális szubjektivitás koncepciójában. Világos, hogy az a felismerés, melyet a *Kartezianus elmélkedések* szövege úgy fogalmaz meg, hogy „a magam számára apodiktikusan adott ego, melyet egyedülként tételezhetek abszolút apodikticitásában, a priori csakis világtapasztaló ego lehet” (Husserl 2000. 156), többet jelent a tudat és világ közötti ismereti korreláció újrafogalmazásánál. Ez a meggyőződés inkább magának a korrelációs problematikának az elmélyítéséről tanúskodik, amennyiben túlmutat az intencionalitás noétikus-noématikus szerkezetű elemzésén, és olyan tapasztalati regiszterek bevonását követeli meg a fenomenológiai elemzésbe, melyek már nem kizárólag episztemikus vagy episztemológiai jelentőségűek. A világtapasztalat a priori jellegének hangsúlyozása ugyanis nem csupán szorosabbra fűzi tudat és világ viszonyát Husserlnél, de megnyitja az utat a konkrét ego, az interszubjektivitás és az életvilág jelenségeinek értelmezése felé is.<sup>1</sup>

A késői husserli elemzések gyakran a „konkrét szubjektivitás” kitételrel utalnak arra a fenomenológiai terepre, amelyet az új típusú vizsgálódások jelölnek ki. Ez a megfogalmazás azonban korántsem hoz minden tekintetben újat Husserl korábbi meggyőződéseire képest. Változatlan marad például az először az 1913-as *Egy tiszta fenomenológia és fenomenológiai filozófia eszméi* című műben szisztematikusan kifejtett tétel, mely szerint a transzcendentális szubjektivitást *egologikus* tapasztalati szerveződés jellemzi. Hasonlóképpen nem történik válto-

<sup>1</sup> Jegyezzük meg, hogy Gerd Brand már egy 1954-ben megjelent és ma sem idejétmúlt művében világossá tette a husserli fenomenológia ezen összefüggéseit. Vö. Brand 1954.

zás Husserlnek abban a meggyőződésében sem, hogy a világtapasztalat a közvetlen *külső észlelésben* nyer eredeti artikulációt. Mindez pedig összefoglalva azt jelenti – állítja Husserl egy 1920-as években keletkezett kéziratában –, hogy a „konkrét szubjektivitás fogalmát az a két, egymástól elválaszthatatlan tény határozza meg, hogy én létezem, illetve hogy közvetlen tapasztalattal rendelkezem a világról” (Husserl 1973b. 380–381). Más szavakkal, a szubjektivitás „zárt konkrétságát” az „egologikus szféra és a hozzá közvetlenül kapcsolódó világtapasztalat rendszere” alakítja ki (Husserl 1973b. 379).<sup>2</sup> Ugyanebben a kontextusban tér ki Husserl arra is, hogy a konkrét szubjektivitás terepe kijelöli a fenomenológiai értelemben vett „eredetiség” hatókörét. Míg az ego és annak sajátos tapasztalatai kapcsán „eredendőségről” (*Uroriginalität*) kell beszélnünk, a közvetlen tárgyi jellegű világtapasztalat az „eredetiség” (*Originalität*) dimenzióját nyitja meg (Husserl 1973b. 386–390). Ez a két meghatározás arra utal, hogy a szubjektivitás konkrét élete és világa mindenütt az intuitív közvetlenség formájában áll fenn. Együttesen pedig azt a *fenomenális léptékváltást* teszik hozzáférhetővé, amelynek alapján az ego és a világ adottságának eltérése a maga tisztán fenomenológiai minőségében tárul fel.

A konkrét szubjektivitás egyszerre egologikus és világszerű tapasztalati hatókörének kijelölése ugyanakkor olyan meghatározásokat is játékba hoz, melyek többé-kevésbé hiányoznak a korábbi husserli elemzésekből. Ilyen például azoknak a jelenségeknek a köre, melyeket Husserl összefoglalóan a „világiasítás” (*Verweltlichung*) terminusával jelöl. A világiasítás egyrészt utal arra a folyamatra, melyben a transzcendentális ego egy sajátos appercepció eredményeképpen önmagát mint világi *pszichofizikai létezőt* fogja fel.<sup>3</sup> Másrészt hangsúlyosan ide tartoznak azok a konstitúciós szintek, melyek az *interszubjektivitás* különböző tapasztalati regisztereihez kötődnek, legyen szó akár az ego és a társszobjektum viszonyáról, akár a közösségiség kialakulásának magasabb formáiról (vö. Husserl 2006. 172–173). Harmadrészt, s talán legfontosabb jellemzőként, a világiasítás arra a módra vonatkozik Husserlnél, ahogyan a konkrét szubjektivitás a világban a maga *testi mivoltában* megjelenik. Ez utóbbi mód jelentősége abban áll, hogy a testiség ugyanúgy alapvetően belejátszik az *ego* pszichofizikai appercepciójába, mint – a természet konstitúcióján keresztül – az interszubjektív viszonyok kialakításába. Röviden tehát elmondható, hogy a transzcendentális szubjektivitásra vonatkozó késői husserli elemzések egyik radikális újdonságát az adja, hogy a konkrét szubjektivitás *testileg megalapozott* tapasztalati szerkezetként kerül bemutatásra. Az alábbiakban ennek a tapasztalati szerkezetnek az elemzésére vállalkozunk.

<sup>2</sup> Lásd még Husserl 1973c. 105.

<sup>3</sup> Lásd ehhez Husserl 2000. 115–116.



## II. A TEST MINT EREDETI MEGJELENÉSSZERKEZET

Fontos látni, hogy azok a fenomenológiai vizsgálódások, melyek a test fenoménjét teszik kitüntetett tárgyukká, legyen szó annak eleven vagy fizikai megjelenési formáiról, korántsem csupán a transzcendentális szubjektivitás „konkrétságára” vonatkozó reflexió eredményei Husserlnél. A test kérdése valójában már jóval azelőtt felmerül a husserli elemzésekben, hogy a szubjektivitás doktrínája transzcendentális formát öltene. Az 1907-es *Dolog és tér* címmel tartott egyetemi előadások már alaposan elemzik a test konstitutív szerepét a külső észlelés vonatkozásában és ezen keresztül a tértapasztalat kialakításában (vö. Husserl 1973d. 44–57. 82–83. §). A néhány évvel későbbi, *A fenomenológia alapproblémái* című kurzus pedig fontos megállapításokat tesz a tudat és a test viszonyáról a természet általában vett konstitúciója szempontjából (vö. Husserl 1973a. 113–125). Ezeket az elemzéseket azonban mégiscsak az *Egy tiszta fenomenológia és fenomenológiai filozófia eszméi* című, a transzcendentális idealizmus tétele mellett már elkötelezett mű második kötete rendszerezi. Husserl itt nem csupán annak kimutatására vállalkozik, hogy az eleven test (*Leib*), mozgási és tapintási képességénél fogva, a materiális és természeti dolgok tapasztalatában konstitutív „szervként” viselkedik, de arra az alapvető szerepre is rámutat, amelyet a test a „pszichikai realitás” konstitúciójában betölt (vö. Husserl 1952. 55–97 és 143–161). Ennek alapján a test a pszichikum eredeti formájaként kerül bemutatásra, amely a testtapasztalat érzéki jellegében és a test mozgásra való képességében gyökerezik. Megjegyzésre érdemes az a tény is, hogy azokban az 1914 és 1918 között keletkezett szövegekben, ahol Husserl az *Fszmék* első kötetének megjelenése után először teszi kifejezett elemzés tárgyává a „transzcendentális idealizmus” fenomenológiai jelentését, a szubjektivitás testi meghatározottsága újra csak erőteljes hangsúlyt kap.<sup>4</sup>

Ugyanakkor, ha valami világosan megváltozik a kései Husserl testiségre vonatkozó elemzéseiben, vagyis azokban, amelyek a „világ” és a „világiasítás” fenoménjeit közvetlen összefüggésbe hozzák a szubjektivitás elemzésével, akkor az éppenséggel maga a *konstitúciós problematika*, amelyből kiindulva az eleven és fizikai test jelenségei fenomenológiailag elemezhetővé válnak. Ahelyett ugyanis, hogy megelégedne annak kimutatásával, hogy a test és a hozzá sajátosan tartozó „kinesztétikus” intencionalitás miként járul hozzá a külső tárgyak és a természet konstitúciójához, Husserl egyre inkább annak a módnak a feltárását tartja elsődlegesnek, ahogyan a konkrét szubjektivitás saját testi jelenlétéből kiindulva *megalapozza* saját tapasztalati szféráját, a hozzá közvetlenül tartozó környező világot, valamint azokat az intencionális utalásokat, amelyek az interszubjektív

<sup>4</sup> Vö. Husserl 2003. 132–140, 160–166. Újabbban Tengelyi László vizsgálta a világ, a test és az interszubjektivitás aspektusainak szerepét Husserl e szövegeiben arra a megállapításra jutva, hogy belőlük a transzcendentális idealizmus egy új, módszertani értelme bontható ki (Tengelyi 2010).

tapasztalat legalsó rétegét képezik. Egyszóval, ahelyett, hogy a test problematikáját az intencionális objektiváció és a pszicho-fizikai egység terminusaiban értelmezné, a késői husserli elemzések abból a tényből indulnak ki, hogy „a világ oly módon konstituálódik, hogy az én eleve testiesült (*verleiblicht*) módon lép fel a világban” (Husserl 1973c. 287). Vagy másként mondva:

Az énstruktúra és a hozzátartozó konstitutív élet, a megjelenések és velük korrelatív létezők a maguk transzcendentális teljességében, mindez világiasul (*verweltlicht sich*), ami azt jelenti, hogy a transzcendentális szubjektivitás testiesül (*verleiblicht sich*) (Husserl 1973c. 289).

Döntő jelentőségű, hogy ezeknek a fenomenológiai viszonyoknak a feltárásában Husserl mindig szem előtt tartja, hogy a „testiesülés” *adottságmód*, mégpedig egy egyszerre eredendő és eredeti adottságmód. Ez a jellegzetesség pedig arra utal, hogy a test tapasztalata korántsem homogén. Ez utóbbi legalább három különböző, noha egymásra épülő szinten ragadható meg, melyek a *hülétikus*, a *kinesztétikus* és a *külső tárgyak* észlelésére jellemző megjelenési módok szerint artikulálódnak.

Valójában már az *Eszmék* második kötetének elemzései utalnak arra az alapvető tényre, hogy az eleven testiség tapasztalata olyan „érzéki benyomások” (*Empfindnisse*) közegében képződik, amelyek „közvetlen intuitív módon” állnak fenn.<sup>5</sup> Ez a közvetlenség olyan minőségek formájában jelentkezik, mint a fájdalom, a gyönyör, az energikus feszültség és nyugalom, a test valamire való képességének vagy képtelenségének érzése. Husserl ezzel egy olyan „hülétikus alapréteg” jelenlétére mutat rá a testtapasztalásban, amelyen keresztül – mint írja – „az ember teljes tudata bizonyos módon kapcsolódik a testéhez” (Husserl 1952. 153). Ennek az elemzési iránynak a jogosságát későbbi szövegek is megerősítik. Husserl egy 1931-es jegyzetében így fogalmaz: „az eleven test az érzékiségek vagy az érzéki benyomások helye, melyek megjelennek rajta és benne, többé-kevésbé megkülönböztetett formában oszlanak el, helyek szerint, lokalizálva, részben bizonyos módon kiterjedve” (Husserl 1973c. 324). Mindehhez hozzátehetjük, hogy a „lokalitás” ténye arra utal, hogy Husserl szerint a testi érzékiség közvetlenül kötődik az érzékszervekkel való rendelkezés tapasztalatához. Ebben az értelemben az érzékiség e szintjét egy tisztán „affektív” adottságmód eredményének lehet tekinteni, amely „passzív” módon konstituálódik (vö. Yamaguchi 1982. 70–73, 107–117). Ebből viszont azt is megtudhatjuk, hogy a testtapasztalat közvetlen adottsága nem a test pszichikai vagy fizikai egységként való *felfogását* jelenti Husserlnél. Ellenkezőleg: a testi tapasztalat

<sup>5</sup> Husserl 1952. 153. Az *Empfindnisse* kifejezés értelmezéséhez Husserlnél lásd Lévinas 1988. 145–165.

mint eredendő *átélés* a test *átérzésére* utal, amely egyúttal a test legitimebb önmegjelenési formája.

Ez a hülétikus adottságmód ugyanakkor nem az egyetlen olyan megjelenési forma, amely meghatározza az eleven test eredendő konstitúcióját. Általánosan szólva, a passzivitás szerkezete mindig feltételezi olyan aktív funkciók meglétét Husserlnél, amelyek az affektivitást a tapasztalati szerkezet kifejezett részévé emelik. Ebből a szempontból alapvető, hogy Husserl a „mozgásérzékben” vagy „kinesztétikus érzékelésben” ismeri fel azokat a funkciókat, melyek eredendő módon illeszkednek a testi affektivitás jellegéhez. Az *Eszmék* második kötete ennek kapcsán már így fogalmaz: „az érzéki aktusokhoz a kinesztétikus érzetek sorozata járul, melyek motiválják az előbbieket a tőlük való függésükben” (Husserl 1952. 57). A kinesztétikus adatok a test eredendő érzéki konfigurációjáról tanúskodnak, amennyiben a test „szabad mozgásra” való képességének tapasztalati feltételeit jelölik. Ahogy Husserl a *Kartezianus elmékedésekben* megállapítja: „szerveim kinesztézise a »cselekszem« módjában folyik le és a »képes vagyok rá« jegyében áll” (Husserl 2000. 113). A mozgásérzék adottságmódjának eredendő mivolta ugyanakkor azt is jelenti, hogy itt a test mozgásának vagy mozgásra való képességének *belső* érzetéről van szó. Husserl kifejezetten hangsúlyozza, hogy az emberi test csak azért képes mozgásra a fizikai térben, mert az eleven test belülről uralt és képzett mozgásai irányítják (vö. Husserl 1973b. 515–516; 1973c. 278–280). A mozgás ebben az értelemben nem fizikai helyváltoztatást jelent, hanem szubjektív módon tapasztalt készletet, feszültséget és irányítást. Noha a mozgásban a test természetesen megváltoztatja a helyzetét a külső tárgyakhoz képest, a mozgás belső, kinesztétikus érzete olyan szubjektív élmények sorozataként konstituálódik, amelyben aktivitása dacára az én csak egy másik mozgó test analógiája alapján képes saját térbeli mozgását érzékelni (Husserl 1973c. 515–516). Összefoglalva tehát elmondható, hogy Husserl koncepciója szerint az eredendő testi tapasztalat csak azért lehet érzékileg *önaffektív*, mert egyszersmind *öneffektív*, vagyis a képességek és cselekvések immanens érzéke ruházza fel eredendő tapasztalati minőséggel.<sup>6</sup>

A test effektivitása annál is inkább döntő jelentőségű fenomenológiai ismerv Husserl számára, mert a testi mozgások és cselekvések ténylegessége kiküszöbölhetetlenül a külső, fizikai tárgyakkal analóg adottságmód szerkezetét is játékba hozza. Ez a jellegzetesség a test eredeti, vagyis világra vonatkoztatott konstitúciójának kulcsa. Husserl gyakran megjegyzi, hogy a kinesztézisek lehetőségére már bizonyos módon megköveteli, hogy a test „térbeliesült” legyen, azaz olyan térbeli dimenzió összefüggésében jelenjen meg, amelynek adottsága nem

<sup>6</sup> Megjegyezhetjük, hogy egyes későbbi munkajegyzeteiben Husserl oly módon definiálja a „kinesztétikus hülét”, mint amely közvetlenül a testi „ösztönök” konstitúciójának szintjéhez tartozik, s enniben a testiség önaffektív jellegéről árulkodik. Lásd ehhez Yamaguchi 1982. 177 és Depraz 2001. 8–9.

vezethető vissza minden tekintetben a belső testi mozgások jellegére. A test ugyanis a külső észlelésben a fizikai test (*Körper*) módján is adódik, más fizikai testek között elhelyezkedve: „az eleven test olyan módon konstituálódik, hogy testrészekről a térbeli dolgok mintájára szerzünk tapasztalatot, ahogyan azok külső dolgokként megjelennek az észlelésben” (Husserl 1973c. 296). Ez a tény ugyanakkor nem teszi lehetővé Husserl szerint, hogy saját testemre pusztán a fizikai tárgyak módján tekintsek. Mivel nem vagyok képes sem elhagyni, sem csupán kívülről szemlélni a testem, ezért az valójában alkalmatlan arra, hogy a külső tárgyakhoz hasonlóan kizárólag észleleti vetületekben jelenjen meg számomra (Husserl 1973c. 268–270). S ez azt jelenti, hogy eredeti külső megjelenése és más dolgokkal való fizikai kapcsolata dacára a testtapasztalat mindig implicálja az érzéki és kinesztétikus élményszerűség eredendő jelenlétét. Ez az élményszerűség pedig végső soron oda vezet, hogy Husserl szerint az emberi test a világban olyan *dologként* jelenik meg, amely *szubjektív tulajdonságokkal* bír:

‘Testem konstitutív tulajdonságai szerint olyan eredeti módon adódó dolog (*Ding*), amelyet eredeti módon mozgatok, amely valamennyi sajátos „mozgásában” és más átalakulásaiban számomra eredeti módon megmaradó egyiséget képez, s amely az önmegjelenés jellegét hordozza, egy olyan valami szubsztrátumának megjelenéseként, ami ténylegesen, lehetősége szerint (és tényleges lehetőségében) számomra eredeti módon adódik (Husserl 1973c. 567).

### III. A TEST MINT AZ OBJEKTIVÁCIÓ FORRÁSA

Mindezek alapján megérthetjük, hogy Husserl miért és milyen módon hajlik arra, hogy a test komplex adottságmódjában egy olyan konstitúció terepét ismerje fel, amelyből kiindulva nem csupán az eredendő szubjektív tapasztalás bizonyos alapjegyei, de az *objektíváció* eredeti tapasztalati folyamatai is feltárhatók. Amit a test fenomenális konstitúciója ebben a tekintetben látni enged, valójában éppen az, ahogyan az objektíváció primitív formái, mint amilyen a tér vagy a természet konstitúciója, saját intencionális képződésükben egy olyan adottságmódra utalnak vissza, amely a tapasztalás legintimebb szubjektív jegyeit hordozza magán. A test fenomenalitása éppen arra a kérdésre adja meg a választ Husserlnél, hogy a *szubjektívítás eredendő önmegjelenése* miként képes motiválni az *eredeti tárgyi megjelenések* feltételeit. Hiszen világos, hogy az érzékszervekkel való rendelkezés immanens érzése – mint amilyen a látás vagy a tapintás – és a test mozgásérzéke alapján előálló tapasztalatok – mint amilyen az orientáció vagy a közelség/távolság élménye – konstitutív módon játszanak szerepet azoknak a tulajdonságoknak a kialakításában, amelyek végső soron a tárgyi világ tapasztalatának formájában jelentkeznek. Ebben az értelemben Husserl számára a test

egy olyan *fenomenális akumuláció* jelölője, amelyben érzékiség és dologszerűség egyöntetűen és egymást kiegészítve hozza létre a szubjektivitás és a világ konstitúciójának egyes alapelemeit:

Testem tehát egy teljességgel sajátos szubjektív-objektív konfigurációban jelentkezik számomra. Nem pusztán dologként a többi dolog között, magánvalóként (az intuitív magánvaló, az intuitív térbeli dolog, egy tisztán szolipszista módon konstituált magánvaló értelmében), hanem figyelemre méltó és meghatározott egységként, mely ugyan analóg a dolgokkal, de megjelenési módjait tekintve mégis a saját testem. Ez az egység magában hordoz egy fizikai testi tárgyiságot, mely nem egy csapásra konstituálódik az eleven test adottságával együtt, hanem egy olyan specifikus és alulról építkező objektiváció szerint, mely a testnek a többi fizikai dologhoz való teljes asszimilációját követeli meg a fizikai oksági láncolatban (Husserl 1973b. 6).

A fenti idézet alapján aligha csodálkozhatunk, hogy Husserl egy 1921-es szövegében miért az emberi test *fizikai konstitúciójának* kérdését tartja olyan „alapproblémának, melyet saját alapjából kiindulva kell végiggondolni” (Husserl 1973b. 77). E kérdés alapvető jellege ugyanakkor nem csupán arra utal, hogy a pszichofizikai konstitúció problémája kétségtelenül a husserli fenomenológiai testelmélet egyik neuralgikus pontját alkotja.<sup>7</sup> Valójában ebben az időszakban Husserlnek már van bizonyos fokig kialakult válasza erre a kérdésre. Az 1925-ös *Fenomenológiai pszichológia* című kurzus szövege egy helyen például így fogalmaz: „a pszichikai legalsó szintje az, ami szomatológiailag pszichikus, amely közvetlenül testiesül és eleven, és amelynek tapasztalatában egyúttal a fizikait is megtapasztaljuk” (Husserl 1962. 131). A szomatológia problematikáján túl azonban a test tisztán dologi aspektusainak fenomenológiai kérdése azért lehet fontos, mert egy olyan objektiváció lehetőségére világít rá, amelyben egy „tárgyi” dimenzió – nevezetesen a testé – nem pusztán *konstituált* létezőként, de *konstituáló* funkciók birtokosaként jelentkezik az eredeti tapasztalás körében.

Husserl számos alkalommal elemzi a test sajátos fenomenológiai jellegét oly módon, hogy hangsúlyozza annak fizikai tárgyi aspektusait és a dologi rendbe való illeszkedését. Így akár nem volna túlzó a *konstitutív tárgyiság* kategóriájáról beszélni itt, abban az értelemben, hogy a test eredeti térbelisége és mozgása konstitutív módon játszik bele valamennyi más külső tárgy eredeti adottságába. Noha a tudat számára az eleven test nem jelenhet meg pusztán egy külsőleg adott tárgy formájában, szemléleti adottságában mégis van valami radikálisan különböző a tudat élményeinek megjelenési módjához képest. A test nem puszt-

<sup>7</sup> Egyes kommentátorok egyenesen leküzdhetetlenek tekintik az ezen a szinten jelentkező fenomenológiai nehézségeket. Lásd Franck 1981. 97–100 és Dodd 1997. 64–74.

ta tárgy, mégis hiánytalanul a tárgyak közé *illeszkedik*. Ebben a minőségében pedig olyan objektív érintkezési felületet teremt a világgal, melynek közegében a test azon tárgyi jellemzők elsődleges tapasztalati bázisának tekinthető – fizikalitás, helyváltoztatás, kauzalitás –, melyek a természet és a dologszerűség konstitúciójának alapelemeit képezik. Ennyiben pedig Husserl joggal hangsúlyozhatja, hogy „a test objektíven elsődleges mindazzal szemben, ami objektív” (Husserl 1973c. 490).

Fontos látni azonban, hogy a testnek ez az objektívációs primátusa mégsem csupán a test fizikai vagy dologi aspektusainak köszönhető. A test mint „tárgy” képtelen volna más tárgyak tapasztalati feltételeinek a megteremtésére, ha ehhez nem járulnának a szubjektivitás eredendő tapasztalati jegyei. Nem véletlen, hogy Husserl már az *Eszmék* második kötetében is a „szubjektív tárgyiság” kitétellemmel jellemzi a test tapasztalati sajátosságát annak materiális meghatározottságaival szemben (Husserl 1952. 153). Ez pedig végső soron oda vezet, hogy a test objektív kitüntetettséget a testtapasztalat szubjektív tulajdonságai készítenek elő. Más szavakkal, a világban való testi jelenlét csak azért válhat objektívációs tapasztalati kiindulóponttá, mert a testhelyzet és a mozgás csakis az érzékek és a kinesztézisek közvetítésén keresztül kerül összefüggésbe a tárgyakkal. Ennek alapján pedig azt állíthatja Husserl, hogy „saját eredeti szférám minden egyes tárgya oly módon van jelen számomra, hogy fennállásukat saját testem és testi eredetiségem révén birtoklom, vagyis az eredeti kinesztézisek révén, melyeket a látás, hallás stb. útján hozok játékba” (Husserl 1973c. 567). A testnek mint az objektíváció forrásának ezt a szerepét az alábbi, 1921-ből származó szövegrész mindennél inkább világossá teszi:

Valamennyi dolog között a testem a legközelebbi, legközelebbi az észlelésben, az érzésben, az akaratban. A test énemmel és működő énemmel a környező világ minden más tárgyát megelőzően alkot egységet. A test úgyszólván a középpont, a középpontban elhelyezkedő tárgy. Középponti működésében birtoklom, és noha (hozzám képest) itt már egy tárgyról van szó, mégis valamennyi más tárgy és valamennyi tárgyra irányuló funkcióm funkcionális középpontját alkotja: ez az, amiből kiindulva és aminek közvetítése révén minden más tárgyhoz hozzáférek, azokat elsajátítom, szándékaimnak megfeleltetem (Husserl 1973b. 58–59).

Ezáltal viszont arra a sajátosságra is fény derül, hogy ha Husserl a test tárgyi jellegét és objektívítását egy későbbi kéziratában az „objektívítás fundamentumának” nevezheti (Husserl 1973c. 490), akkor ez csak azért lehetséges, mert a test adottságmódja már a szubjektivitás és interszubjektivitás eredeti tapasztalatai számára is alapként szolgál.

## IV. A TEST MINT ALAP

A test adottságmódjainak beható elemzése teljes mértékben indokoltá teszi Husserlnél, hogy a konkrét szubjektivitás fenomenológiai problematikáján belül a test és a hozzá tartozó tapasztalati regiszter centrális helyre kerüljön. Ez a sajátosság pedig annak köszönhető, hogy a test – mint láttuk – egyszerre eredendő és eredeti megjelenési szerkezetével ugyanúgy döntő fenomenológiai motívumokat mutat fel a „belső” szubjektív tapasztalat, mint a „külső” tárgyi világ tapasztalatának vonatkozásában. E fenomenológiai jellemzőket a maguk összességében tekintve azt lehetne mondani, hogy a testtapasztalat három különmemű, noha egymástól korántsem független *megalapozási módot* tesz láthatóvá abban a folyamatban, ahol Husserl szerint a szubjektivitás és a világ viszonya konstituálódik, vagyis ahol ez utóbbiak az észlelésekben és a velük korrelatív tárgyi megjelenésekben eredeti módon artikulálódnak. Ez a három megalapozási mód tulajdonképpen a „világiasítás” három rétegének feleltethető meg, amelyen keresztül a szubjektivitás a maga konkrétságában „fellép” a világban.

Az első megalapozási mód a szubjektivitás eredendő *testi önmegjelenésének* szintjén érhető tetten. Ez a megalapozás azonban csupán részben utal a testtapasztalat közvetlen és érzéki adottságmódjára mint eredendő konstitúciós szintre. Itt ugyanis egyúttal arról az észlelési helyzetről is szó van, amelyet Husserl egy helyen a test „immanens módon megalapozott transzcendenciájának” nevez (Husserl 1973b. 249). E megfogalmazás azt jelenti, hogy a tudat és a test eredeti összetartozásának intuitív tapasztalatán túl a testézés és a mozgásérezék arról is eredeti adottsággal szolgál, hogy az eleven test „megjelenő dolog” (uo). Ezáltal pedig arra a tényre derül fény, hogy a szubjektivitás eredeti megalapozó testi szférája egészen a külső, térbeli és természeti tapasztalatig terjeszkedik. Erre az összetett megalapozásra vonatkoznak Husserl azon 1921-es elemzései is, amelyek a testet „kettős tapasztalat tárgyaként” mutatják be, ahol a kérdés éppen az, hogy a „test önészlelése” mennyiben „alapja (*Fundament*)” egyszerre a „test (*Leib*) appercepciójának” és a „testalkat (*Leibkörper*) appercepciójának”.<sup>8</sup> Világos, hogy ez a kérdés nem kevesebbet érint, mint a test egyszerre működő szubjektivitásként és természeti létezőként való konstitúciójának lehetőségét. Ha ugyanakkor Husserl ebben a szövegében még nem ad végérvényes választ erre a kérdésre, megteszi ezt máshol. A *Formális és transzcendentális logika* nem habozik a testtapasztalat kapcsán az „én pszichofizikai létezőként való önkonstitúciójáról” beszélni oly módon, hogy – hasonlóan a *Kartezianus elmékedések*hez – Husserl a test eredeti konstitúcióját a természet egy olyan primordiális rétegével hozza kapcsolatba, amely „önmagában első”, mert „még nem objektív természet” (Husserl 1974. 246–248; 2000. 112–115). E megfogalmazások pedig

<sup>8</sup> Uo. 57. A *Leibkörper* kifejezés „testalkat”-ként való fordításához lásd a *Kartezianus elmékedések* magyar kiadásához fűzött megjegyzéseimet: Takács 2001.

arra engednek következtetni, hogy a test önmegjelenésének ebben a dimenziójában nem kevesebb, mint a szubjektivitás és a világ korrelációjának legelemibb formája alapozódik meg.

A test fenomenalitása azonban egy további értelemben is megalapozónak nevezhető. Ez a megalapozás az *objektívációnak* azzal a rétegével áll összefüggésben, amelyről az előző fejezetben már beszéltünk. Itt Husserl abból a tényből indul ki, hogy a test saját kinesztétikus szerkezetén keresztül nem csupán utat talál a világ dologi és természeti létezőihez, de az eredeti tapasztalat szempontjából egyenesen kondicionálja azokat. Ebben az értelemben már az *Felmérés* második kötete is úgy határozza meg a kinesztétikus adottságmódot, mint „ősalapot (*Urgrund*)”, melynek révén „az én környező világának valamennyi reális tárgya kapcsolatban áll a testtel” (Husserl 1952. 56). A test kinesztétikusan objektíváló tapasztalata éppen azt teszi lehetővé, hogy a világ mint tárgyi horizont a szubjektivitáshoz konkrétan hozzátartozó dimenzióként *jelenhessen meg*. Mert a mozgás – s erről már volt szó – az eleven testtapasztalat számára nem annyira helyváltoztatást, hanem a tárgyi megjelenésekhez nélkülözhetetlen közelség, távolság és orientáció konstitúcióját jelenti. Amikor tehát Husserl a „test objektivitását minden objektívitás fundamentumának (*Fundament*)” minősíti, akkor egy alapvető konstitúciós viszonyt ír le, az eredeti adottságmódok logikáját követve:

A test objektivitása az objektívitás fundamentuma. Az objektívitás az előzetesség és utólagosság rendjébe illeszkedik, mely maga is a primordialis ontológiai rendjén alapul a konstitúció funkciójából származó rendként. A test korábbi a testi valósághoz képest, amely a testen kívül létezik. A testiség pedig, noha az objektív világ meghatározottságainak csupán egy rétegét képezi, korábbi a meghatározottságok azon rétegéhez képest, melyet objektív szellemnek nevezünk (Husserl 1973c. 490).

A testtapasztalat megalapozó jellegének harmadik, egyben utolsó elemét az *interszubjektivitás* szintjén fedezhetjük fel. Husserl fenomenológiai elemzései itt legtöbbször abból a megállapításból indulnak ki, mely szerint az *ego* és az *alter ego* közötti viszony alapelemei a testi konstitúció szintjén keletkeznek. A *Kartézianus elméletek* elemzései ebből a szempontból példaértékűek, és világossá teszik, hogy a másik legelőször is külső testi jelenlétével képez tapasztalati tartalmat számomra. Csakis az ezen a tapasztalaton alapuló *hasonlósági appercepció* révén juthatok el ahhoz a következtetéshez, hogy a velem szemben megjelenő létező a *társszubjektum* tulajdonságaival bír (vö. Husserl 2000. 125–130). Ha tehát a másik szubjektum tapasztalata valamiképpen az *eredeti alapítás* (*Urstiftung*) jegyével rendelkezik számomra, akkor ez Husserl szerint kizárólag a testi és a testalkati megjelenés szerkezetéből bontható ki (vö. Husserl 2000. 127–129). Itt pedig azok a sajátosságok az iránymutatók, amelyek magát az eredeti testi megjelenést jellemzik. Az eleven test külső tárgyként való megjelenési képes-



sege eredeti módon motiválja a hasonlósági appercepciót, olyannyira, hogy Husserl egy jegyzetében egyenesen arról beszél, hogy „a test térbeli dologként való konstitúciója [...] alapja (*Fundament*) egy másik idegen test közvetett appercepciójának” (Husserl 1973b. 516. 1. jegyzet). Vagy még pontosabban:

Azt kell hogy mondjuk, hogy egy idegen test észlelése lényegét tekintve egy eredeti értelmezésen keresztül történő észlelés. Ez az eredendőség a saját primordiális testiségemre való kiküszöbölhetetlen utalásban alapozódik meg (*gegriündet*), amelyben eredeti élménnyel rendelkezem egy olyan szubjektív lét testiségéről, amely dologként jelenik meg (Husserl 1956. 63).

Ennek alapján világossá válhat annak a husserli állításnak az indítéka is, amely szerint „a közösség formájában konstituálódó mozzanatok közül az első a természet, amely valamennyi interszubjektív közösségiség alapja (*Fundament*)” (Husserl 2000. 137). Ha Husserl elemzéseiben ugyanakkor olyan fenomenológiai megoldásokat is keres a közösség konstitúciójának kérdésére, amelyek nem követik minden tekintetben a *Kartezianus elmékedésekben* kifejtett „analogikus appercepció” doktrínáját, a testtapasztalat akkor is alapvető kiindulópont marad a számára. Legyen szó, akár a „monásközösség” tapasztalati feltételeinek elemzéséről, akár az „ösztön” és a hozzá tartozó sajátos intencionalitás feltárásáról, a test fenomenalitása mindezen koncepciókban olyan alapot nyújt Husserl számára, amely az interszubjektív konstitúció viszonyait továbbra is az eredeti megjelenések tényéből és szerkezetéből engedi kibontani.<sup>9</sup>

Husserl fenomenológiai testelméletének szempontjából nem tűnik aggályosnak egy olyan megállapítás elfogadása, mely szerint „az emberi testre vonatkozó tapasztalatunk bizonyos módon a valóságra vonatkozó tudásunk alapját képezi, s ezáltal metafizikai fogalmainknak is alapul szolgál” (Mezei 2004. 162). Hiszen a konkrét szubjektivitás természetére vonatkozó husserli elemzések nem csupán a tudat és a világ transzcendentális korrelációjának újfajta leírására tesznek kísérletet. Legalább ennyire fontos, hogy a testiség fenomenológiai elemzését Husserl egyrészt a természet, másrészt az interszubjektivitás konstitúciójának a kérdéséig terjeszti ki. Ez a megközelítés pedig amellet, hogy a testtapasztalat alapvető szerepét hangsúlyozza mindkét említett konstitúciós módban, egyúttal folyamatosan a pszichofizikai tapasztalat és létmód problémáját tárgyalja. Ezekben a pontokon, anélkül, hogy le kellene mondaniuk a fenomenológiai módszer által biztosított talajról, a husserli elemzések világos metafizikai jelentőségre tesznek szert. Mert a testtapasztalat felbontása érzéki és természeti összetevőkre már

<sup>9</sup> A monadikus tapasztalat és testtapasztalat megalapozó viszonyához lásd Husserl 1973b. 261–272 és Husserl 1973c. 634–641. Az ösztönintencionalitás testi megalapozottságához pedig lásd Husserl 2004. 123–126 és értelmező írásomat: Takács 2004. 120–122.

csírájában tartalmazza a tudat fakticitásának és a világ transzcendentális értelmének téziséét.<sup>10</sup> A husserli „metafizika” azonban az eredeti tapasztalati adottságmódok elemzésén nyugszik. S ez az eljárás gyakorlatát és doktrinális kidolgozottságát tekintve a mai napig fenomenológiai követőkre vagy konkurenszerekre vár.

## IRODALOM

- Brand, Gerd 1954. *Ich, Welt und Zeit. Nach unveröffentlichten Manuskripten Edmund Husserls.* Den Haag, M. Nijhoff.
- Depraz, Natalie 2001. *Lucidité du corps. De l'empirisme transcendantal en phénoménologie.* Dordrecht–Boston–London, Kluwer.
- Dodd, James 1997. *Idealism and Corporeity. An Essay on the Problem of the Body in Husserl's Phenomenology.* Dordrecht–Boston–London, Kluwer.
- Franck, Didier 1981. *Chair et corps. Sur la phénoménologie de Husserl.* Paris, Minuit.
- Husserl, Edmund 1952. *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie, Zweites Buch: Phänomenologische Untersuchungen zur Konstitution.* (Husserliana IV.) Den Haag, M. Nijhoff.
- Husserl, Edmund 1956. *Erste Philosophie (1923–1924), Zweiter Teil: Theorie der phänomenologischen Reduktion.* (Husserliana VIII.) Den Haag, M. Nijhoff.
- Husserl, Edmund 1962. *Phänomenologische Psychologie. Vorlesungen 1925.* (Husserliana IX.) Den Haag, M. Nijhoff.
- Husserl, Edmund 1973a. *Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlass. Erste Teil (1905–1920).* (Husserliana XIII.) Den Haag, M. Nijhoff.
- Husserl, Edmund 1973b. *Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlaß. Zweiter Teil (1921–1928).* (Husserliana XIV.) Den Haag, M. Nijhoff.
- Husserl, Edmund 1973c. *Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlaß. Dritter Teil (1929–1935).* (Husserliana XV.) Den Haag, M. Nijhoff.
- Husserl, Edmund 1973d. *Ding und Raum. Vorlesungen 1907.* (Husserliana XVI.) Den Haag, M. Nijhoff.
- Husserl, Edmund 1974. *Formale und Transzendente Logik. Versuch einer Kritik der logischen Vernunft.* (Husserliana XVII.) Den Haag, M. Nijhoff.
- Husserl, Edmund 2000. *Kartezianus elmélkedések.* Ford. Mezei Balázs. Budapest, Atlantisz.
- Husserl, Edmund 2003. *Transzendentaler Idealismus. Texte aus dem Nachlass (1908–1921).* (Husserliana XXXVI.) Dordrecht–Boston–London, Kluwer.
- Husserl, Edmund 2004. Egyetemes teleológia. Ford. Kukla Krisztián – Takács Ádám. *Vulgo.* 5/1. 123–126.
- Husserl, Edmund 2006. *Späte Texte über Zeitkonstitution (1929–1934). Die G-Manuskripte.* (Husserliana Materialien, Band 7.) Dordrecht–Boston–London, Kluwer.
- Lévinas, Emmanuel 1988. Intentionnalité et sensation. In uő *En découvrant l'existence avec Husserl et Heidegger.* Paris, Vrin. 145–165.
- Mezei Balázs 2004. Az emberi test modális elmélete. In Boros Gábor – Ullmann Tamás (szerk.) *A szubjektum problémája. Kortárs fenomenológiai tanulmányok.* Budapest, Tudástársadalom Alapítvány. 157–210.

<sup>10</sup> Husserl metafizikai kérdésfeltevéseinek értelmezéséhez lásd Schwendtner 2008. 230–243.

- Schwendtner Tibor 2008. *Husserl és Heidegger. Egy filozófiai összecsapás analízise*. Budapest, L'Harmattan.
- Takács Ádám 2001. Az evidens és az örökkévaló (E. Husserl: *Kartezianus elmékedések*). *BUKSZ.* 13/3. 221–222.
- Takács Ádám 2004. Bevezetés Edmund Husserl *Egyetemes teleológia* című szövegéhez. *Vulgo.* 5/1. 120–122.
- Tengelyi László 2010. A fenomenológia módszertani idealizmusa. *Magyar Filozófiai Szemle.* 54/1. 9–22.
- Yamaguchi, Ichiro 1982. *Passive Synthesis und Intersubjektivität bei Edmund Husserl*. Den Haag, M. Nijhoff.

## A feminista elmélet és a (női) test

### Joó Mária: A feminista elmélet és a (női) test

Írásommal a feminista elmélet és a magyar filozófiai diskurzus közötti kapcsolatot szeretném intenzívebbé tenni: ezért bemutatok az adott témakörben néhány jelentős irányzatot és szerzőt.<sup>1</sup> A testről szóló feminista írások mintegy külön diskurzussá álltak össze: úgynevezett test-feminizmus (*corporeal feminism*) néven is szokás rájuk hivatkozni. Ezen belül két eltérő felfogást mutatok be: az egyiket inkább fenomenológiaiként jellemezném, képviselői közül főleg Iris Youngra szorítok, a másikat inkább posztstrukturalistának tekintem (a test konstruktivista-diskurzív felfogása értelmében), itt a kevésbé ismert Susan Bordót emelem ki Butler mellett. A két felfogás inkább hangsúlybeli eltéréseket mutat, mivel abban egyetértenek, hogy a testben minden egyszerre természeti és kulturális, ezért egyik-másik képviselőjük általam használt besorolása vitatható (Bordo például inkább a két szélső pólus közötti elméleti pozíciót képvisel). Ha azzal akarnám jellemezni a két csoportot, hogy milyen klasszikus forrásokból merítettek, akkor azt mondhatjuk, hogy az előbbit Merleau-Ponty és Beauvoir filozófiája, a másikat Foucault és a posztstrukturalisták inspirálták.

<sup>1</sup> A két diskurzus közti kapcsolat meglehetősen egyoldalú – bár országonként változó mértékben: a feminista elmélet művelői többnyire jóval jártasabbak a másikban, lévén az a domináns diskurzus. Mindazonáltal az utóbbi művelői időnként sommásan elítélő véleményt fogalmaznak meg, informátlanságuk ellenére. Erre egy francia példát hoznék: Luc Ferry az ökológiai gondolkodást bemutató könyvében kitér az ökofeminizmusra (III. A különbözőség dicsérete), és mivel nem látja át a mára már számos különálló területre tagolódott feminista diskurzusokat, azoknak tulajdonítja a saját eligazodási képtelenségét: „összefüggéstelen gondolathalmazt” lát, amelyben – számára – „összegabalyodik a pszichoanalízis, a szociológia és a politikatudomány”. Valójában ő tekint egy pszichoanalitikus értelmezést történeti igazságnak, és ekként visszautasítja. Jogosan. (Ferry 1992/1994. 255 skk.)

## I. FEMINISTA ALAPOZÓ

Mielőtt tulajdonképpen témámra térnék, talán nem árt röviden összefoglalnom – mivel nem tételezhetem föl, hogy köztudott – a Nő helyét a filozófiában.<sup>2</sup> A nő volt a férfi mellett a Másik, mássága a testében, sőt – Beauvoirt parafrázálva (Beauvoir 1969. 7, 11) – a méhében gyökerezett (Szent Tamás szerint). A tradicionális felfogás szerint a nőt létfeltételei bezárják testébe, s így szokás volt a szexualitással azonosítani. Világos, hogy mindez a férfi látószögéből állítható, aki magát tudatnak, abszolútumnak, saját testét pedig normális kapocsnak tekintette tudata és a külvilág között, míg a női testet gátnak, börtönnek látta.<sup>3</sup> A női másság történetileg társadalmi másodrendűséget, alárendeltséget jelentett, amit többnyire testi okokra vezettek vissza (gyengébb fizikum, szülés). A férfi magát az Ember lényegi képviselőjének tekintette, a nőt pedig a nem lényegi másiknak. Az egyoldalúság nem szándékos ármány, nem dominanciaigény, illetve nem kizárólag a tesztoszteron hatásának köszönhető – legalábbis a filozófiai elméletek magyarázata szerint. Sajátos tudományfilozófiai probléma jelenik meg abban, hogy miközben a filozófiai elméletek az általánosság objektív igényével jöttek létre, mégis részleges igazságot produkáltak, mely az emberiség felét kevésbé vette figyelembe.<sup>4</sup> Ezért a feminista elméletek elsődlegesen kritikai elméletekként születtek, és ma is ez az egyik lényeges szerepük. Természetesen nem privilégiumuk a kritikai aspektus: gyakran párhuzamosak a kortárs, a filozófiai diszciplínán belüli kritikai áramlatokkal (Nietzsche, Foucault vagy a posztstrukturalisták), amiből kölcsönhatások, esetenként vádak, elsőbbségi viták adódnak, például a feminista elméletek és a posztmodern áramlatok között (Séllei 2006).

A test tehát szimbolikus értelemben női, míg az ész férfi konnotációjú kategória volt a filozófiában, alapul szolgálhatott az ész- és férfiközpontúság (Derrida kifejezésével élve *fallogocentrizmus*) kritikájára.

<sup>2</sup> A magyarul feminista filozófia címszó alatt olvasható könyvek vagy túl szélesen fogják föl a kérdést, vagy csak a témában jártasak számára érthetőek. Hell Judit műve (2006) egy *gender studies*hoz való interdiszciplináris-történeti bevezetésnek alkalmas, míg Herta Naglé (2006) filozófusoknak való, amennyiben ismerik a hivatkozott műveket. Annemarie Pieper könyve (2004) ugyan feminista etikát ígér, de inkább a feminista filozófiába való bevezetesként használható. Leginkább Rosi Braidotti (2007) tanulmányát ajánlanám. A további szakirodalmat lásd: Joó 2004, főleg 64–66; a bibliográfiát pedig: 79–85.

<sup>3</sup> Vö. Beauvoir 1969. bevezetés, főként 7–11.

<sup>4</sup> A tudományfilozófiai elv vagy probléma a tudás részlegessége, történeti-társadalmi-nemi szituációhoz kötöttsége, melyet a tudásszociológiától a feminista nézőpontelméletig (*stand-point theory*) és Donna Harawayig sokan elemeztek.

## II. CORPOREAL FEMINISM

A testtel foglalkozó feminista diskurzus képviselői közé tartoznak Iris Young, Susan Bordo, Elizabeth Grosz, valamint az ausztráliai feminista iskola más képviselői (Rosalyn Diprose, Moira Gatens), és még sokan mások. Judith Butler a kilencvenes évek elején megjelent első két könyvét is ide sorolhatjuk, ezek magyarul is olvashatóak (Butler 2005, Butler 2006). Számos gyűjteményes kötet készült Susan Rubin Suleiman korai összeállításától (1985) egy 1999-es *Feminist Theory and the Body* című reprezentatív válogatáson át (Price–Shildrick 1999)<sup>5</sup> a legújabb, nálunk is hozzáférhető, *The Body – A Reader* című könyvig (Fraser–Greco 2005).<sup>6</sup> A gazdag anyagból praktikus szempontból két területet emelek ki, melyek iránt nálunk különös érdeklődést remélhetek: a fenomenológus Merleau-Ponty, valamint Foucault hatását tükröző feminista kutatásokat,<sup>7</sup> melyek a női testi tapasztalat szempontjából kritikusan újraértékeltek és kiegészítették, továbbfejlesztették az univerzálisnak tartott elméleteket.

A két általam választott területet, a fenomenológiai feminizmust és a Foucault-követő vagy posztmodern feministákat egy az utóbbi évtizedben kibontakozott vita kapcsolja össze. Ez a nemi identitás testes és konstruktivista felfogása között zajlik, amely más, testi jelekben azonosítható identitásra, így a fajra (*race*) is érvényes. Nevezhetjük konstruktivizmus-vitának is.<sup>8</sup> A test neme, a bőr színe velünk születik, de megszabja-e maradandó módon sorsunkat, jellemünket, tulajdonságainkat, vagy inkább a hozzájuk kapcsolódó történelmi-társadalmi normák a döntőek, s ezek konstruálják a nemi, faji identitást, sőt magát a testet is – női testként, fekete bőrű testként, persze a mi közreműködésünkkel? Beauvoir tette fel először a kérdést a nőkre vonatkozóan: Mi a nő, mit jelent nőnek lenni, milyen tapasztalatokat, érzéseket, sorsot? Az egzisztenciális fenomenológia az észlelés, a testi tapasztalat jelenségéből, valamint a szubjektumnak a mindenkori szituáció keretei között adott alapvető szabadságából indul ki, vagyis összekapcsolja a belső és a külső szempontot. „Senki nem születik nőnek [vagy férfinak – tehetjük hozzá], hanem azzá válik” (Beauvoir 1969, 197). A sokat idézett beauvoiri mondat egyesek szerint a későbbi *biológiai nem / társadalmi nem* megkülönböztetés alapjává vált, bár vitatható, hogy jogosan-e. Mint-

<sup>5</sup> Csaknem 500 oldal, sajnos rövidített szövegekkel.

<sup>6</sup> Welton 1998 gyűjteménye azt példázza, hogy egy mai reprezentatív válogatás feminista kutatásokat, szerzőket is kell tartalmazzon (a 16 szerzőből legkevesebb 8 feminista kutatóként ismert, Bordótól 3 írás, Youngtól szintén 3, Butlertól 2).

<sup>7</sup> Price (1999) válogatásában nem szerepel Iris Young, akit én – nem egyedülként – különösen jelentősnek tartok, bár lehet, hogy az azóta eltelt több mint egy évtizedben vált világossá az ő elméleti súlyát, miután egy kötetben (Young 2005) újból megjelentek a női testi tapasztalatról szóló korábbi és újabb írásai. Ezzel szemben Welton (1998) három írását is közli.

<sup>8</sup> Vö. Wanick–Stoller 2001 és Welton 1998 *Contested constructions* című részével. Konstitúciós felfogásnak is nevezem. Lásd később Butler felfogása kritikájánál, ahol Nagl (2006) szóhasználatát veszem át.

egy három-négy évtizeddel az alapító atyák/anyák (Beauvoir) után a feminista elméletben a probléma kezelésére sokáig hasznosnak bizonyult megkülönböztetés (*sex/gender*, biológiai nem / társadalmi nem) problematikussá vált. Judith Butler híressé vált 1990-es könyvének címe erre utal: *A problémás nem* (Butler 1990/2006). A kilencvenes évektől Foucault, Butler és a dekonstrukció terjedő népszerűsége nyomán egyre inkább a nemi identitás, a szexualitás diszkurzív konstruáltságának felfogása kezdett dominálni: a test mint írható felület, szöveg, illetve a normák testi beíródása (*bodily inscriptions*), reprezentációja vált a kutatás fő irányává. Szállóigévé vált Butler első könyvének egyik konklúziója: a *sex* már mindig is *gender* volt. Vagyis nincs elkülöníthető *sex*, illetve *gender* (sőt csak *gender* van), és a kettő között nincsen oksági kapcsolat (tudniillik, hogy adott nemi szervekkel születvén mintegy automatikusan nővé/férfivá válánk). Ez a belátás első pillantásra nem áll távol a fenomenológiától, vagyis attól az alapfelfogástól, hogy test és kultúra, test és lélek nem választható el egymástól. Azonban úgy tűnik, ettől elég távol került a test egyértelműen társadalmi konstruáltságát valló posztmodern, konstruktivista feminizmus, amely erősen támaszkodik a pszichoanalitikus elméletekre. A tapasztalat és elemzési kategóriáinak történetileg változó voltát bemutató Joan Wallach Scott, tekintélyes feminista történész egyetlen magyarul megjelent, alapvetőnek számító írásában, amely a társadalmi nem definícióját adja, homályban hagyja a test jelentőségét, miközben Foucault-ra és Butlerre hivatkozik.<sup>9</sup> A fenomenológiai alapokhoz jobban kötődő feminista kutatások az eltűnő testet igyekeztek visszahelyezni a centrumba, a megélt testet, a testi, érzelmekkel telített prediszkurzív tapasztalatot próbálták megragadni. Ebben főleg Beauvoirra és Merleau-Pontyra támaszkodhattak, ahogy ezt Iris Marion Young korai, nyolcvanas években írott műveiben ki is mondja. Elizabeth Grosz, Moira Gatens és Christine Battersby is jelentős műveket alkotottak, mindhárman az úgynevezett ausztráliai feminista körhöz tartoznak.<sup>10</sup> Ide sorolható a megélnkülő Beauvoir-kutatás is, hiszen számos kutató szerint nála is inkább Merleau-Ponty hatása dominál, nem annyira Sartre-é.<sup>11</sup> Sonia Kruks kiemelkedő teljesítményt nyújtott a tapasztalatnak a diskurzusok szorításából történő visszahódításában; ő is az egzisztenciális fenomenológiától indult, és Beauvoir-kutatóként is jelentős.<sup>12</sup> Elizabeth Grosz Lacan pszichoanalitikus elméletét átértelmezve, Nietzscheből, Foucault-ból és Deleuze-ből, főként azonban Merleau-Pontyból merít, különösen a *Megélt testek: fenomenológia és a hús* című fejezetben (Grosz 1994). Groszhoz képest Iris

<sup>9</sup> A *gender* fogalom kimerítő definícióját adó cikkében: A társadalmi nem mint a történeti elemzés hasznos kategóriája (Scott 2001. 126–161, főként 144–147). Talán relevánsabb lenne *Experience* című cikkét elemezni: Butler–Scott 1992. 22–40.

<sup>10</sup> Gatens 1996; Battersby 1998.

<sup>11</sup> Vö. Heinämaa 2003; Kruks 2001; magam is ezt állítom.

<sup>12</sup> Kruks 2001; főként a *Going beyond Discourse. Feminism, Phenomenology and „Women’s Experience”* című fejezet, 131–153.

Young érdeklődése inkább a politikai filozófia területére sorolható, ha létezik a politikai fenomenológia kategóriája, akkor ő pontosan ezt művelte: az elnyomás változatos arcait bemutató írása nálunk is ismertté vált.<sup>13</sup>

### 1. Merleau-Ponty nyomában: Iris Marion Young<sup>14</sup>

Mint közismert, Merleau-Ponty a testre vonatkozó fejtegetéseiben nem megy tovább annak a megállapításánál, hogy a test nemi/szexuális létező, a szexualitás koextenzív a létezéssel, minden mozzanatát áthatja a nem.<sup>15</sup> A nemi különbség megjelenik mind az észlelés, mind a tapasztalat, a kifejezés, a mozgás stb. során. Elsőként Merleau-Ponty barátját és egyben első méltatóját, Simone de Beauvoirt inspirálta alkalmazásra és továbbgondolásra *Az észlelés fenomenológiája*, melynek legismertebb bizonytsága az 1949-ben megjelent *A második nem*. Mással foglalkoztam Beauvoirrall, most kettőjük együttes hatását, továbbfejlesztését mutatom be Iris Youngnál.<sup>16</sup> Csak utalok rá, hogy Judith Butler, akiről a továbbiakban lesz szó, szintén írt egy korai kritikus elemzést *Az észlelés fenomenológiájáról* (Butler 1989),<sup>17</sup> valamint Luce Irigaray francia pszichoanalitikus feminista is, aki a késői Merleau-Ponty *khiasmus*-fogalmát fedezi fel női-feminista szempontból (Irigaray 1993. 151). Őket manapság a posztmodern feminizmus megjelöléssel illetik, bár néha mindkettőjüket megtalálhatjuk a pszichoanalitikus feminizmus címszó alatt is.<sup>18</sup> Butler filozófiai indulása egyértelműen a francia egzisztenciális fenomenológiához, a francia Hegel-recepcióhoz kötődik (Butler 1987) – van, aki egy mai egzisztencialistát lát benne.

Iris Marion Young kiemelkedő feminista filozófus volt (néhány éve halt meg). A politikai filozófia területén is jelentős műveket alkotott, most azonban a női testi tapasztalat autentikus feltérképezőjeként mutatom be.<sup>19</sup> A tárgyban a

<sup>13</sup> *Five Faces of Oppression*. In Young 1990b. A szöveg (elég gyenge) magyar fordítása egy konferenciakiadványban jelent meg.

<sup>14</sup> Merleau-Ponty hatása tekintetében Young mellett Grosz 1994, Bigwood 1991 és egy általam nem látott gyűteményes kötet (Olkowski 2006) érdemel említést. Vö. még a 20. jegyzettel.

<sup>15</sup> Állításomat elsősorban *Az észlelés fenomenológiájára* korlátozom, mivel késői műveit kevésbé jól ismerem. Kiváló magyar munka Vermes 2006, amely a nemi különbség tematizálásának a küszöbére ért, de – Merleau-Pontyhoz híven – nem lépett át rajta.

<sup>16</sup> Joó 2003; 2005.

<sup>17</sup> Talán kicsit felületesre sikerült ez az írás, ahogy Waldenfels (2005. 249. jegyzet) felröta neki: mintha *Az észlelés fenomenológiájának* csak egy fejezetére támaszkodna, és így megalapozatlan vádakat fogalmaz meg, például a biologizmusét.

<sup>18</sup> Így a Csabai–Erős-féle válogatásban (1997) Butler filozófus, Irigaray pszichoanalitikus, de mindketten foglalkoznak a másik szakterületével. Butler mindkét magyarul olvasható könyvében több fejezet elemzi Freud, Lacan és Irigaray műveit, míg Irigaray (1993) sajátos filozófiatörténetet ad.

<sup>19</sup> Welton 1998 három írását közli *Situated Bodies* cím alatt: a dobásról, a terhességről szólót és a dobáshoz írt húsz évvel későbbi kommentárt.



nyolcvanas és kilencvenes évektől kezdve írt tanulmányai egy kötetbe gyűjtve mutatják igazán Young innovatív teljesítményét, melyek a sajátos női tapasztalat értelmezésében kreatívan felhasználják más gondolkodók (például Roland Barthes) és más feminista diskurzusok (például a pszichoanalitikus Irigaray, Kristeva) eredményeit is. (Sokarcú filozófus volt, talán egyoldalúan hangsúlyozom fenomenológiai vonásait.)<sup>20</sup> Egyik első és híressé vált írása a *Lányos dobás. A női testi viselkedés, mozgás és tériség*,<sup>21</sup> mely egy kötet címadó írása lett. Ebben olyan további élmények, megélt tapasztalatok sorakoznak, mint a terhesség testi léte (*pregnant embodiment*), a mellek tapasztalata: a látvány és az érzés, a ruhák hordása (Young 1990). A következő évtizedben Young politikai filozófiai írásai mellett folytatta ezt a fenomenológiai projektet, sorra véve a ház és az otthon, a menstruáció, az öregek testi gondozása női térénümát, gyakran konfrontálva ezeket e témák első női-feminista leírásával, melyet Beauvoirnál találunk.<sup>22</sup> Közös jellemzője az írásoknak, hogy elkerülik az egyoldalú negatív ábrázolást, a női létet nem kizárólag a patriarchális kultúra fogságában, a férfiak nézőpontjának kiszolgáltatott tárgyként, áldozatként látják, hanem azt is megmutatják, hogy ugyanazt a jelenséget új, pozitív módon is lehetséges átélni. A nő nemcsak alávetett, hanem „visszahódíthatja” ruháit, terhessége, otthoni munkája értelmét. (Ebben az ambiguitásban Young megint csak egyetérthet Beauvoirról, még ha egyes tapasztalatok nyomasztó beauvoiri ábrázolásával vitatkozik is, így például a terhesség és a szülés esetében [Young 2005]). A leginkább negatív a labdadobás nőies módjának leírásában, ebben ugyanis szerinte a sajátos női modalitás ineffektív, a végtagok használata, mozgatása a testet körülvevő térben korlátozott, a női dobás a férfiéval összehasonlítva célszerűtlen, tehát idéltlen, vagyis rosszabb. Ez a női testi mozgásstílus és leírása mára többé-kevésbé elvesztette aktualitását, talán csak nyomokban lelhető fel. Az írás a hetvenes évek végén született, az ötvenes–hatvanas években felnőtt női generációkban még jellemző testi modalitás lehetett a zárt, finomkodó mozgás, mivel a lányok testnevelésére kevesebb súlyt fektettek, mint a fiúkéra, és a nőies ruházkodás normája kötelező szoknyaviseletet írt elő. Azóta viszont a nőkre is vonatkozik a sportos testideál, amelyet a lazább ruházat, a nadrágviselés elterjedtsége kísér. Ezt a

<sup>20</sup> Inkább csak ebben a korai írásában hivatkozik Merleau-Pontyra, a későbbiekben más hatásokat (Foucault stb.) is asszimilált gondolkodásába, ezért külön kifejtést érdemelne Merleau-Ponty hatásának ki-, illetve bemutatása. Vö. Preston 1996 kritikáját, illetve Bigwood 1991. Merleau-Ponty növekvő jelentőségét mutatja egy külön kötet, amely feminista interpretációival foglalkozik (Olkowski 2006).

<sup>21</sup> *Throwing Like a Girl. A Phenomenology of Feminine Body Comportment, Motility and Spatiality.* In Allen–Young 1989, és külön saját kötetében címadóként más tanulmányokkal együtt: Young 1990. Utolsó kötete (Young 2005) ezekből négyet, valamint újabbakat is tartalmaz; a rá jellemző éleslátással az öregkor és a testi ápolásra szorulás intézményes formáit mutatja be (a magánélethez való jog hiányát).

<sup>22</sup> Ezek közül a jelenségek közül a menstruáció, a terhesség, a szülés meglehetősen nyomasztó élményekként jelennek meg Beauvoirnál, vö. Joó 2003. 42.

másféle női testi habitust azonos és koedukált nevelés is előmozdítja. A két nem mozgásának modalitása nem vált ugyan teljesen azonossá, de sokban közeledtek. Nyomokban még ma is találkozunk a női motorikus készségek lekicsinylő sztereotípiáival (például női autóvezetés), és nyilván valóságával is, s még él a különbség normája férfiak és nők ruházatában, célszerűség és divat tekintetében. A divatos ruházatkodás a nők mozgását inkább korlátozza, mint a férfiakét (a magas sarkú cipő és szűk miniszoknya viselőjének nem teszi lehetővé teste szabad, célirányos használatát például labdahajításkor). A lányos dobás esetét azért sem tartották a Merleau-Ponty-kutatók jó példának, mivel nők esetében ez nem habituális mozgás, így implicit módon a férfi labdadobás lett a mércé, a jó dobás.<sup>23</sup> Ezt a kifogást az idő változó perspektívája okafogyottá tette, ma már a labdadobás is lehet női habituális viselkedés (és ekként effektív, jó, még ha más stílusú is).<sup>24</sup>

A többi sajátosan női tapasztalat közül a ruhák hordását emelem ki Young *Nők, visszahódítjuk ruháinkat* című írása kapcsán. A szerző egy reklámfotóból indul ki, melynek szlogenje arra buzdít, hogy képzeljük el magunkat egy gypjűkabátban. A képen egy utcai felvételen egy modellt látunk, aki bő, puha esésű kabátba burkolózik, mögötte egy férfi áll, aki nézi. Young leírja, hogy elképzeli magát ebben a kabátban, amely puha, meleg – szinte látja, érzi –, és azt is érzi, hogy ugyanakkor egy férfi nézi. Tetszik magának, mert tetszik egy férfinak? A közfelfogás szerint a nők azért szépítik magukat, hogy a férfiaknak tessenek. A feminista elmélet szerint a nők a férfiak szemével látják magukat szépnek vagy csúnyának, passzív tárgyai, illetve kivetítései a férfi vágyának, élvezetének, s az ő nőkről alkotott képeinek akarnak megfelelni. A feminista elméletben a *male gaze* (férfitekintet) az ezen a területen végzett elemzéseknek összefoglaló megjelölése.<sup>25</sup> A képen látható modell szép, sötét hajú, titokzatos nő (napszemüveget visel). Az őt néző nő (a szerző) azonosul a modellel, elképzeli magát abban a helyzetben, abban a kabátban, olyan szépnek és titokzatosnak, mint amaz – ez azonban az ő saját élvezete és öröme, amelyet a képzelet révén él át. Így a fantázia segítségével nemcsak a férfi tetszésének, élvezetének a tárgya, hanem saját maga érez saját örömet. Ez a divatlapok, filmek, reklámok sikerének titka – lehetőséget ad különböző helyzetek, identitások (sportos, falusi, romantikus, nagyvilági stb. nő) átélésére, játékra, szabadságra. Young *A divat rendszerére* (Barthes 1999) hivatkozva a divat retorikáját, szemiotikai rendszerét idézi, amelyben az élvezetet a tudatosan nem valódi, elképzelt világ szabadsága okozza. Ezáltal a nők visszahódítják ruháikat, amelyek

<sup>23</sup> Preston (1996. 167–186) azt állítja, hogy a példa minden nem habituális mozgásra vonatkozik, vagyis nem speciálisan női. Egy férfinál a megfelelő időtlen mozgás/cselekvés például pelenkacsere esetében figyelhető meg.

<sup>24</sup> A testi habitusok nemileg különböző voltának Bourdieu szentelt később nagy figyelmet, bár ő éppen ezek stabilitását, alig változó voltát emeli ki, és ebben látja a patriarchális kultúra mai újratermelődésének (öntudatlan) módját (Bourdieu 2000).

<sup>25</sup> Néhány klasszikus írás magyarul is megjelent a témában, például Mulvey 2000, és ugyanott de Lauretis és Tania Modleski írásai.

eredetileg a férfiak tetszésének elnyerése céljából készültek, amelyek önkifejezési lehetőséget adnak számukra, tudniillik fantazmatikus azonosulást tesznek lehetővé elképzelt sorsokkal, szerepekkel.<sup>26</sup>

Young a tapintás – és kevésbé a látás, a nézés – elsődlegességét és fontosságát hangsúlyozza a ruhák okozta érzéki tapasztalat során, bár a fantázia egyben vizuális is (Young 2005. 69). Luce Irigaray-val együtt a sajátos női érzékiség, élvezet közegének a tapintást tartja, míg a látás inkább a férfi élvezet számára elsődleges. A tapintás közelséget, érintést feltételez, míg a látáshoz távolság, elkülönülés kell. A testi érintés, tapintás jelentőségét szintén a késői Merleau-Pontynál találjuk meg, ezt Irigaray is felfedezte és elemezte (Irigaray 1993). Saját értelmezésének egyik újdonsága, hogy ebben a tapasztalatban és a róla való beszédben látja azt a közös élményt, amely a nők közötti kulturális különbségek ellenére kapcsolatot teremt nő és nő között.

Ugyanakkor a feminista kritika elsődleges teljesítménye annak az anonim hatalmi kényszernek az elemzése, amelynek a nők többnyire önként meg akarnak felelni, női identitásuk megalkotásában ugyanis természetesnek érzik, interiorizálják a nőiség normáit, amelyek elsődlegesen a kinézetre, a testre vonatkoznak. Foucault szexualitástörténetének feminista elsajátítása korán elkezdődött, ezt a normalizálási kényszert az egyik első szerző, Sandra Lee Bartky szépség- és divatkomplexusnak nevezte.<sup>27</sup> A képeknek elnyomó hatalmuk is van, a tökéletes, szép, karsú női testek látványa a saját test tökéletlenségének érzését kelti: a nők többségének a saját testéhez való viszonya negatív, elégedetlen vele, és eközben állandóan mintegy kívülről látja magát, tárgyként, tükörben.<sup>28</sup> A női tükör a férfiak vágyának ideális tárgyához méri a valódi testet, ennek a női testideálnak az utóbbi fél évszázadban lényegi vonása a soványság/vékonyosság/karsúság. Az állandó fogyókúra, a test sanyargatása éhezéssel jellemzően női viselkedés, bár nem kizárólag. A női test jellegzetes tünete a mai korban az evészavar, annak különböző súlyossági formái: anorexia, bulimia. Freud korában még a hisztéria volt a jellegzetesen női betegség, amely mára eltűnt. Susan Bordo az elsők között elemezte a sovány test jelentését, illetve az anorexia üzenetét ma már klasszikusnak számító írásaiban.<sup>29</sup>

<sup>26</sup> Természetesen a divat, a női ruházzkodás komplex jelenség, megjelenik benne a kultúra története, gazdagsága, sok elemzési lehetőséget ad a *cultural studies*, *women studies* témájaként, sőt létezik *fashion studies* is.

<sup>27</sup> Lásd Bartky 1992; 1990. A szépségmítoszok a tömegkultúra jelenségeire vonatkozó elemzését lásd Wolf 2000.

<sup>28</sup> Foucault *Panoptikum*-képe sokkal elementárisabb női tapasztalat, mint férfi: a nőket kislánykoruktól arra nevelik, hogy állandóan ügyeljenek a külsejükre. Ebben különleges szerepet játszik a tükör, amelyet Irigaray 1974/1985 elemez. Ez az érdekes könyv többek között Platón barlanghasonlatát és Freud *Nőiség* című írását tárgyalja.

<sup>29</sup> Lásd Bordo 1993 írásait: *Anorexia Nervosa: Psychopathology and the Crystallization of Culture; Reading the Slender Body: The Body and the Reproduction of Femininity*. E kötet egyik írása magyarul is megjelent, lásd Bordo 2006.

## 2. Bordo, a karcsúság kulturális normája és tapasztalata

Mielőtt a címben jelzett témáról írnék, érdemes visszautalni a test-feminizmus általam említett két csoportja közötti éles elhatárolás nehézségeire nemcsak Butler, hanem Bordo esetében is. Bordónál is egyértelmű Foucault hatása: a sovány testek kulturális, diszkurzív konstitúcióját – ebben az értelemben a „posztmodern testet” – elemzi (Bordo 1993. 215–301), tehát a másik csoportba is tartozhatna. Bordo rámutat egy tényre, amelyet sok feminista elméletíró is elfelejt, miközben Foucault fogalmainak használhatóságát méltányolja: a hatvanas–hetvenes évek nyugati nőmozgalmi tudatosították a nők tömegeiben kulturális elnyomottságukat, és ezáltal váltak érzékennyé Foucault egy évtizeddel később angolul is megjelenő műveinek hatására. Az amerikai feminizmus teljesítménye Bordo szerint a testi praxisok politikai jelentésének tudatosítása, amelynek jelzőszerű megfogalmazása, hogy minden személyes/magánjellegű dolog egyben politikai jelentéssel is bír („*The private is political*”), és mi lenne személyesebb, mint a test? Foucault kései írásaiban néhol utal a feminista mozgalmakra, de az elméletírókat kevésbé ismerhette.

Bordo írásai közül az anorexia jelentésének elemzésével foglalkozó szöveget emelem ki, amely a nyolcvanas évek közepén íródott, és az azóta eltelt több mint két évtized igazolta lényegbevágó voltát. A fogyókúrázás tömeges komplexussá vált, és már alsó tagozatos gyerekek is fogyni akarnak. Az utóbbi években kezdtek e veszélyes jelenség megváltoztatására irányuló bátortalan erőfeszítések megjelenni, legalábbis a médiabeli bejelentések szintjén (tudniillik hogy véget kell vetni a soványságuideál terrorjának) – egy divatlap címlapján már egy-egy kövér modell is megjelenik (*Elle* magazin, 2010. május).

Az *anorexia nervosa* csak pszichopatológias formája az állandó sikertelen fogyókúrák tömegeket érintő testi-lelki praxisának, amelyben a mai kultúra egy problémája kristályosodott ki. Természetesen nem maga a jelenség új, csak a tömeges elterjedése – már Binswanger (1958) is írt egy anorexiás nőbetegéről. Lényegében az európai kultúra egyik tradicionális vonása képezi a jelenség alapját, a test megvetése, a vele való helyes bánásmód, amely a megfegyvelés, kordában tartás, leigázás erkölcsi parancsában fogalmazódik meg Platóntól Descartes-ig és a keresztény életelvekig. A test e tradícióban hitvány szolga az ész uralma alatt. Bordo a nyolcvanas évek amerikai kultúrájának a közegében értelmezi ezt a sokdimenziós evési rendellenességet, amelyről akkor és azóta is számos szakmunka és népszerű tudományos ismeretterjesztő írás jelent meg, az utóbbi évtizedben már magyarul is. A témával kapcsolatos amerikai szerzőket és a betegek vallomásait nem ismertetem, mivel számunkra a fogalmi-filozófiai keret az érdekes. Minden érintett életében más-más története és sokdimenziós oka van e táplálkozási rendellenességnek, amely szélsőséges formában halált okoz, egy ponton túl ugyanis visszafordíthatatlanná válik a patológikus fogyás. Több fajtája is van, mindezekről számos internetes oldalon lehet információt

találni. Ami a sokféle jelentésben mégis közös, azt Bordo három tengely köré csoportosítja: ezek a dualizmus-, a kontroll- és a nemi/hatalmi tengely.

A dualizmus-tengely a test–lélek dualizmusra utal, amely Platón, Szent Ágoston és Descartes filozófiájában ismert módon meghatározta az európai gondolkodást a testről. A mai testsanyargató diéták, kemény testedzések, a test tiltakozásának, fájdalmának, éhségének figyelmen kívül hagyása a régi aszketikus erkölcsi ideálok folyamánként lehetnek ma is ilyen erősek, éppen mert mögöttük van egy több mint kétezer éves tradíció, amely lenézte, megvetette a test kívánságait és örömeit, fájdalmait pedig elviselendőnek, netán egyenesen üdvösnek tartotta. (A test a lélek börtöne, a vágy – az evés, szexuális élvezet iránt – az ész veszedelmes ellensége, hogy csak Platónról idézzek.) Az anorexiás az éhséggel harcol, a testén lévő zsírpárnák ellen küzd, sikeresen egyre soványabbá válva. Célja az éhség testi érzésének megsemmisítése.

A kontroll-tengely a test sikeres uralásának élménye, az akaraterő diadala: nemcsak a fogyás, hanem egy edzett test kialakításának/előállításának (izmok, hajlékonyság stb.) a céljából is. A testi tapasztalat a rendszeres kemény testedzéssel járó önelgyőzés (például napi 15 km futás, 1-2 órás *working out* egy fitness centerben), a test ellenállásának leküzdése (az izmok és ízületek terhelése, a tüdő- és szívkapacitás növelése a lehetőségek határáig). A mai fitnessmánia gyakran beteges formát ölt, kényszerré válik és a tökéletes testi kontroll élményét adja, (extrém fájdalomtűrés, a mindenhatóság fantáziája). A testből mára a kultúra fogyasztási tárgyat, fétist csinált: fiatalsága, rugalmassága, simasága és vékonysága minden erőfeszítést, fájdalmat és pénzt megér.<sup>30</sup> A gazdagoknak a plasztikai sebészet szolgáltatja a test elképzelt formára való átalakításának lehetőségét a vele járó rizikó ismertetése nélkül.<sup>31</sup> (A művészi önkifejezés test adta lehetőségei, a női body art, a performanszok feminista kritikai és kreatív törekvései külön kifejtést érdemelnének, amelyre itt nincs módom.)<sup>32</sup> Az ilyen ideálisra fabrikált test érték – szimbolikusan is, „gazdájának” elismerést szerez –, valójában azonban szolgálként, alakítható tárgyként, eszközként bánnak vele. A tökéletes testnek elsősorban nőneműnek kell lennie – bár nem kizárólag. Más jelentése is van a női testnek: nemesak önmaga, hanem férfitpartnere szimbolikus tőkét is képviseli (a fiatal, vonzó női test státuszszimbólum). Ezzel elérkeztünk a harmadik tengelyhez.

A nemi/hatalmi tengely: A test neme a hatalmi viszonyok elsődleges jelölője (Scott 2001. 146). Egy nő teste, külseje összehasonlíthatatlanul fontosabb társa-

<sup>30</sup> Vö. Baudrillard 1998.

<sup>31</sup> Heves viták övezik a feminista elméletben, sokan elutasítják (pl. Andrea Dworkin), de vannak (pl. Kathy Davies, lásd in Price 1999. 454), akik a női önrendelkezés, szabadság részeként helyeslik, ha a nő maga választja „szabadon” a kötelező szépségideál követését. Lásd még Morgan *Sculpted Bodies* című írását (Morgan 1998).

<sup>32</sup> A magyar Drozdik Orsolya művészete feltétlenül említést érdemel, a számos hetvenes-nyolevanes évekbeli amerikai avantgard művésznő mellett. A saját testén végzett plasztikai műtéteket nyilvános performanszá alakító Orlanról lásd Fraser–Greco 2005. 43.

dalmi megítélése, státusza szempontjából, mint egy férfié. A tradicionális férfi–nő hierarchiával összefüggésben a nő saját testtapasztalata, testképe többnyire a hiányosságok, hibák meglétére, észlelésére koncentrálnak. Szociológiai felmérések igazolták a női önértékelés, a nő saját testével való elégedettségének túlnyomóan negatív voltát, szemben a férfiak testi önértékelésének túlnyomóan elfogadó jellegével. A nők többsége kövérnek tartja magát, illetve egyes testrészeit – annak ellenére, hogy esetleg a különféle testindexek szerint nem is számít annak, és annak ellenére, hogy mások soványnak tartják. A női vonzerő elképzelhetetlen zsírpárnákkal – például egy színésznő esetében, míg egy középkori férfiszínésznél elnézik, ha deréktájban kicsit megvastagodott. A soványság normáját nemesak önmaga elé állítja egy nő, hanem gyakran férfitartnere kéri számon rajta, és a női ikonfigurák (modellek, színésznők) formáival hasonlítja őt össze – s ítéli kevésbé vonzóknak, szépeknek. Mások értékelése felülírja a saját test képét – az anorexiás is kövérnek látja magát. Ő azonban sikeresen legyőzte éhségét és uralja testét. A tökéletes önkontroll élménye erőt ad, a nő az akaratérő diadalát éli meg – és a beteg még a klinikai stádiumban is azt hiszi, hogy uralja a testét, és bármikor abba tudja hagyni a koplalást, ha akarja. (Ez sajnos többnyire illúzió, gyakran már klinikai úton sem lehetséges az anyagsere helyreállítása.) A feminista elemzések összefüggést találtak a női testi formák, folyamatok kamaszkori jelentkezése és az anorexia között. A serdülő lány tudattalanul tiltakozik a nővé válás és az azzal járó kulturális normák ellen. A fiúsan vékony anorexiás test biztonságot ad, és tagadja a nőiséget, a kontrollképesség férfias fantazmatikus identifikációt tesz lehetővé – tudattalanul. Régi kulturális mintákat fedezhetünk fel e mögött a modern kórkép mögött. Az anorexiás azt a nőiséget tagadja és kontrollálja saját testében, amelyhez a féktelen vágyak képzete tapadt (szexuális étvágy, evés vágya, érzelmek iránti vágy stb.). A nő mindig „túl sok”, maga a természet elementáris ereje – ezért is szorul rá a férfi irányítására, akinek segítségével megtanulhatja önmagát uralni (Rousseau Zsófiája és Emilje az ideális pár példája). Ennek a kulturális mintának (archetípusnak) megtestesítői Éva, Pandóra, a boszorkány, a varázslónő, a vámpír, a *femme fatale*. Pszichoanalitikus értelmezés szerint a női hatalom, erő archetípusai a férfiak nőktől való félelmének kivetítései – és ezt interiorizálja büntudatként a nő, aki hibásnak, túlzásnak, „túl soknak” érzi magát.<sup>33</sup> Gay (1984. 197–201) összefüggést lát egy-egy ilyen alakzat megjelenése és a nők időnként észlelhető társadalmi megerősödése között, mint például a *femme fatale* és a 19. századi szüfraszett-mozgalom esetében.

A test konkrét, megélt anyagiságát Bordo inkább hangsúlyozza, mint a következőkben bemutatásra kerülő, hangsúlyjaiban eltérő álláspontot képviselő Butler. A két szerzőt mások megítélése szerint is érdemes egymás mellé állítani (Hekman 1998. 91).

<sup>33</sup> Eichenbaum–Orbach (1983) és Dinnerstein (1976. 59–62) szerint ezt a fantazmát a korai gyermekkor hatalomteli anyafigurájától való függőség élménye alapozza meg.

### III. DISZKURZÍVE ELŐÁLLÍTOTT NEMISÉG – JUDITH BUTLER KONSTRUKCIÓS TESTFOGALMA

A test kulturális kényszerek, normák formálta felfogásának és megélésének egyik szélsőséges konstitúciós felfogása Butler két, kilencvenes évek elején írt művében jelent meg. Erősebb formában az elsőben, a *Problémás nemben*, és némileg enyhített mértékben a *Jelentős testekben*, amelyben a kritikai reflexiók hatására módosította, újragondolta egyes állításait. (Megjegyzéseim csak erre a két könyvre vonatkoznak, nem célolok Butler átfogó értékelése, még kevésbé a hozzá kapcsolt *queer*-elméleté.)

Butler a mai feminista elmélet sztárja, annak amerikai értelmében. Emellett jelentős filozófus, immár két évtizedes alkotói pályával, aktívan politizáló értelmiségi, művei a politikai filozófia területére is kiterjednek (az identitáspolitikai problémáiról, például a gyűlöletbeszédéről is írt könyvet). Igen vitatott gondolkodó, nem utolsósorban bonyolult, esetenként homályos stílusa miatt: egy alkalommal a „Rossz stílus verseny” (*Bad Writing Contest*) első helyezettje volt.

A *Problémás nem – Feminizmus és az identitás felforgatása* 1990-ben jelent meg. Célja az volt, hogy megkérdőjelezze a női identitás stabilitását és esszencializáló felfogását, amely a kötelező érvényű heteroszexuális mátrix mentén keletkezik. Még a feminizmus politikai céljai, a nők közös érdekei sem igazolják, illetve teszük szükségessé az egységes női identitást – bár ilyen célok megfogalmazhatók, de más elméleti alapon, nem fundacionalista módon is lehetséges a nők politikai összefogása (Butler 1997). A feminista elméletben használt *sex/gender* megkülönböztetésen keresztül mutatja be, ahogy – szerinte – a *gender* társadalmi-kulturális kategóriája hozza létre a természetes vagy biológiai nemet (*sex*), amennyiben ezek a diszkurzív gyakorlatok konstituálják a nemi testet. Ezek a nemi diskurzusok jelölik ki a tapasztalat határait, egy adott kultúrán belül lehetséges módjait, érthető formáit. A nemi test egy folyamat eredménye, illetve állandóan ismétlődő gyakorlatok során materializálódik:<sup>34</sup> maga a test „megtestesülése” ezeknek a gyakorlatoknak. A foucault-i alapgondolat (regulatív normák) felismerhető, ezen kívül Butler továbbgondolja és átviszi a performatív nyelvi aktusok austinai gondolatát a testi viselkedésre, amely révén állandóan előadjuk, sőt létrehozuk női vagy férfi (heteroszexuális) voltunkat. Ezáltal Butler dekonstruálja a természetes vagy biológiai nem fogalmát: a test a kulturális inskripció számára szubsztrátum, írható felületként funkcionál. A nem tehát performatív testi aktusok nem szűnő és ismétlődő sora, cselekvéselőállítás, de nem valami eleve adottnak a kifejezése.<sup>35</sup> A felforgató aktusok azok, melyek leleplezik valaminek a csináltságát, például egy transzvesztita a nő/férfi imitációjával felfedi a nem utánzó és

<sup>34</sup> Az ismételhetőség fogalmát a magyar fordítás meghagyta a latin-os angolos „citalhatóság” formában (Butler 2005).

<sup>35</sup> A fogalmi tisztázáshoz különösen használható Butler 1993/2005 előszava (11–35), és még inkább a *Performativitás mint citalhatóság* című fejezet (25 skk.; én „ismételhetőség” formában

így esetleges jellegét (Butler 2006. 234). A híres-hírhedt parodikus performansz ötlete ehhez a gondolathoz kapcsolódik – később ezen kritikusi sokat élcelődtek.<sup>36</sup> A nem performativitásának gondolata a következő évtizedekben nagy hatással volt a feminista irodalomtudományban (Bollobás 2007). A kötet differenciált elemzéseket tartalmaz a pszichoanalízis (Freud, Lacan, Kristeva), Monique Wittig, Foucault (*Herculine Barbin*), Joan Riviere műveiről. Butler három évvel későbbi könyvében, amely az angolban beszédesebb címmel jelent meg (*Bodies That Matter...*),<sup>37</sup> folytatja a kérdezést a testről mint anyagról. Mit jelent a test anyagisága? És a szexus<sup>38</sup> anyagisága (materialitása)? Honnan van neme a testnek? Butler azt az elterjedt felfogást kérdőjelezi meg, amely az anyagra mint prediszkurzív entitásra hivatkozik. Szerinte már maga az anyag fogalma „telített a szexus (*sex*) és a szexualitás beszédmódjaival, amelyek behatárolják és korlátozzák az 'anyag' szó használatát” (Butler 2005. 41, 239). Annak bebizonyítását tűzte ki célul a műben, hogy „konstitutív történelme során ezt a redukálhatatlan anyagiságot egy problematikus nemi mátrix hozza létre”. Ha ez sikerül – írja –, „akkor az is belátható, hogy az a diszkurzív gyakorlat, mely az anyagnak redukálhatatlan jelleget ad, egyben ontologizálja és rögzíti is ezt a nemi mátrixot” (uo.). Az anyagiság fogalma eleve női, „a nemi különbség drámájának színtere”.<sup>39</sup> Butler a nemi mátrix genealógiáját vizsgálja meg. Természetesen a görögöknél kezdi, de a mai diskurzusok keretében: Arisztotelésznél Foucault-val és Platónnal, a *Timaios* *khóra*-fogalmát újragondolva Kristevával, Derridával és Irigaray-vel vet számot.<sup>40</sup> Nyilván már a korábban kifejtettek is sokakban felidéztek anyag és forma viszonyának arisztotelészi felfogását, amely az élőlények keletkezésének elképzelésében az anyaghoz mint önmagában semmilyen minősítéssel nem rendelkező szubsztrátumhoz a nőt rendeli hozzá, a formához a férfit, amely formát ad az anyagnak, illetve a születendő emberpalántának. Ez az arisztotelészi gondolat közel áll a formáknak (az ideáknak) és a *khórának* a kozmosz keletkezésében játszott szerepéhez Platón *Timaios* szájában. Butler alapvető értelmezési keretként használ egy általa revidált lacani és freudi elméletet.<sup>41</sup> Lacan elmélete, fogalmi apparátusa a két nőtanítvány, Julia Kristeva és Luce

használtam a szót). A szerző itt az előző könyvét ért kritikákra is kitér, melyek nyilván segítettek a fogalmi tisztázásban.

<sup>36</sup> Nem egészen méltányos módon Martha Nussbaum (1999) is, aki Butlert a paródia professzoraként aposztrofálta – elsősorban azért, mert a posztmodern divattól féltette a klasszikus liberális feminista politika népszerűségét.

<sup>37</sup> Butler 2005. Az angol *matter* szó „anyag” és „számít, fontos” jelentéssel is rendelkezik. (Az indoeurópai nyelvekben a *matéria*, *mater* szócsalád jelentésmezője sok inspirációval szolgált az anyai testről szóló elméletekhez.)

<sup>38</sup> Ezt a megoldást választották a fordítók az angol *sexre*.

<sup>39</sup> Platón *Timaiosza*: Butler 2005. 59.

<sup>40</sup> Butler Irigaray-vel vitatkozó *khóra*-értelmezéséről lásd Irigaray-ról és a platóni erósról szóló írásomat: *A platóni erősz pszichoanalízisben* (publikálatlan kézirat).

<sup>41</sup> Főleg a *Képzletbeli azonosulás és a Szexus felvétele* című fejezet: Butler 2005. 97–121.



Irigaray gondolati megértéséhez is alapvető, velük Butler helyenként saját értelmezését állítja szembe.<sup>42</sup>

E rövid összefoglalás keretében csak egyetlen lényegi kritikai megjegyzést tennék a performativitás mint materializálódás felfogása kapcsán, ahogy az e két könyvben megjelenik.<sup>43</sup> A testet diszkurzív hatások effektusaként felfogni igen problematikus, mivel a nyelvi performativitás nyelvfilozófiai tézisént ontológiaivá tágítja, a test konstitúciója folyamatának fogalma ezáltal az „előállítás” szinonimájává lesz. A diszkurzív gyakorlatok, a szavak hoznák létre a testet, illetve az anyagot? Ez nagyon emlékeztet a Szó teremtő hatalmának vallási felfogására („Kezdetben volt az Ige”). A nemi testre alkalmazva pedig azt sugallja, hogy szabadon dönthetünk, milyen nemet, testet jelenítünk meg. Butler a kritikus ellenvetésekre válaszolva tagadja, hogy szerinte a test tisztán nyelvi természetű (Butler 2006. 74),<sup>44</sup> de az előző könyv megfogalmazásai teret adtak ennek az értelemnek. Ezzel a kritikai megjegyzéssel nem állok egyedül, például Herta Nagl, a feminista filozófia egyik európai tekintélye is hasonlóan látja (Nagl 2006. 55–57).<sup>45</sup> Úgy vélem, hogy a vitatott gondolat a *Jelentős testek*ben már csak egy lehetséges pszichoanalitikus, szemiotikai értelmezés státuszát kapta. Ezt egy hermeneutikai elméleti keretben elhelyezve Butler felfogása kevésbé problematikus, de egy szigorúbb fenomenológiai keretben még mindig kifogásolható marad. Mindazonáltal ez nem teszi érvénytelenné az általános filozófiai kritikát.<sup>46</sup>

Úgy vélem, hogy Butler testfogalmánál Elizabeth Grosz „megélt test”-interpretációja elméletileg sokkal tarthatóbb, mert megőrzi a nemileg tematizált test szintjén is azt az ambivalenciát, amely ma a posztmodern testelméletek kihívása idejében Merleau-Ponty egyik legfontosabb örökségének tűnik (Grosz 1994. 190). Butlerrel azért is lehet jogosan összehasonlítani, mert mindketten nagymértékben élnek a pszichoanalitikus elméletek értelmezési keretével, bár Groszt a *Volatile Bodies* (Grosz 1994) alapján inkább a Merleau-Ponty követők közé sorolnám (Deleuze és Foucault nem kevésbé hatott rá).

<sup>42</sup> Különösen a *Leszbikus falosz* és a *Morfológiai imaginárius* című fejezet: Butler 2005. 65–97.

<sup>43</sup> Sara Salih írt egy informatív bevezetést Butlerről (Salih 2003); a róla szóló cikkeek száma óriási, afféle kötelező feminista penzumává vált, csak néhány jelentős szerzőt említek később lábjegyzetben.

<sup>44</sup> Itt már csak azt állítja, hogy „a nyelv nem áll ellentétben az anyagisággal”, a nyelv anyagiságáról beszél és a késői Merleau-Ponty „világ teste” fogalmát ajánlja analógiaként. A nyelvi és testi morfológia egyik lehetséges kapcsolatát Lacan és Kristeva pszichoanalitikus elmélete segítségével vázolja fel az adott fejezetben (a híres *Tükörstádium* című Lacan-írásról van szó).

<sup>45</sup> Illetve a *Sex/gender. Hogyan zárhatnánk le egy hosszúra nyúlt vitát?* című fejezet egésze (Nagl 2006. 49–71).

<sup>46</sup> Butler népszerűsége miatt inkább a kritikus értelmezők közül neveznék meg néhányat: az itt említett szerzők közül Bordo (1993, 289. skk), a klasszikus fenomenológia oldaláról Sara Heinámaa és Regula Giuliani, a feminista elméletben pedig Sonia Kruks, Seyla Benhabib és Toril Moi.

Összegzésül a test konstitúciós felfogása körüli vita keretéhez térek vissza. A Butlerrel jelzett elméleti pozíció a nemű esszencializmus vádjá következtében a *genders* (sokféle, tetszőleges és rögzít[het]etlen) álláspontjához, *gender*-szkepticismushoz (Bordo 2006), vagyis a testi szubjektivitás partikularitásához, a heteronormativitás értelmében véve nem nélküli, sokféle testekhez vezetett. Így többen célszerűnek látják a *gender* (társadalmi nem) kategória feladását, és helyette a megélt test fogalmát használni, amely az egzisztenciális fenomenológia értelmében mindig szituált testet jelent, vagyis Beauvoir és Merleau-Ponty együttes örökségét hordozza (Moi 1999. 3–121). A megélt test fogalmát a fenomenológiai indítatású szerzők mind elfogadják, a testetlen, racionális szubjektum fogalma helyett ez feminista szempontból is termékenynek tűnik. Ennek elfogadása azonban nem helyettesítheti, vagy teszi fölöslegessé a *gender* dualizmusát. Bemutatott szerzőink közül Bordo fejt ki azokat a társadalmi és filozófiai érveket, melyek a *gender* kategóriájának szükségessége mellett szólnak – csak kettőt említek: a filozófiai tradíció bináris oppozíciói „kidobásának” elméleti apóriája, valamint a férfiak és nők közötti tovább élő egyenlőtlenségek kritikai tematizálása, a politikai cél kitűzésének szükségessége (Bordo 2006. 111, 118). Iris Young is ezen az állásponton van egyik utolsó írásában, amelyben Toril Moi álláspontjára reflektál: *Lived Body vs. Gender, Reflections on Social Structure and Subjectivity* (Young 2005. 12). Elisabeth Grosz könyvét pedig azért emelném ki újból, mert ő dolgozta ki, mintegy kipreparálva Merleau-Ponty műveit a feminista alkalmazások számára, a *lived body* (megélt test) fogalmát, ezáltal megnyitva az utat egy feminista fenomenológia felé.<sup>47</sup> Grosz jelentősen hozzájárult a nemileg (is) különböző testek különböző tapasztalatainak megértéséhez, és kiemelkedő filozófiai teljesítményként még Lacan patriarchális elméletének is képes volt feminista olvasatát adni, valamint egyéb posztmodern kihívásokkal is megbirkózik.<sup>48</sup>

## IRODALOM

- Allen, Jeffner – Iris Marion Young (szerk.) 1989. *The Thinking Muse. Feminism and Modern French Philosophy*. Bloomington, Indiana University Press.
- Barthes, Roland 1999. *A divat rendszere*. Ford. Mihancsik Zsófia. Budapest, Helikon.
- Bartky, Sandra Lee 1990. *Femininity and Domination. Studies in the Phenomenology of Oppression*. New York, Routledge.
- Bartky, Sandra Lee 1992. Foucault – a nőiség és a patriarchális hatalom modernizációja. Ford. Keresztes György. *Magyar Filozófiai Szemle*. 36/3–4. 434–445.
- Battersby, Christine 1998. *Phenomenal Woman: Feminist Metaphysics and the Patterns of Identity*. New York, Routledge.

<sup>47</sup> Lásd Embree–Fisher 2001, különösen Debra Bergoffen és Linda M. Alcoff cikkeit.

<sup>48</sup> Elisabeth Grosz 1990. *Jacques Lacan – A Feminist Introduction*. London, Routledge. Ennek egy részlete olvasható magyarul Csabai–Erős 1997-ben. Még két másik írása is olvasható Séllei 2006-ban arról a kihívásról, amelyet Derrida és Foucault jelentett a feminizmus számára.

- Baudrillard, Jean 1998. The Finest Consumer Object: Body. In Fraser–Greco 2005. 191–205.
- Beauvoir, Simone de 1969/1971. *A második nem*. Ford. Görög Lívía és Somló Vera. Budapest, Gondolat.
- Benhabib, Seyla – Drucilla Cornell – Judith Butler – Nancy Fraser 1995. *Feminist Contentions. A Philosophical Exchange*. London, Routledge.
- Bigwood, Carol 1991. Renaturalizing the Body (with the help of Merleau-Ponty). In Welton 1998. 99–115. Eredeti megjelenés: *Hypatia*. 1991. 6/3. 54–73.
- Binswanger, Ludwig 1958. The Case of Ellen West. In Rollo May (szerk.) *Existence*. New York, Simon and Schuster.
- Bollobás Enikő 2007. Performansz és performativitás – a női, meleg és nem-fehér szubjektumok nagy előadásai az irodalomban. In Séllei Nóra (szerk.) *A nő mint szubjektum, a női szubjektum*. Debrecen, Kossuth Egyetemi Kiadó. 17–58. (Vö. Bollobás nagydoktori értekezése a performativitásról. Budapest, MTA, kézirat 2009.)
- Bordo, Susan 1993. *Unbearable Weight. Feminism, Western Culture and the Body*. Berkeley, University of California Press.
- Bordo, Susan 2006. Feminizmus, posztmodernizmus, genderszkepticizmus. Ford. Séllei Nóra. In Séllei 2006. 93–130.
- Bourdieu, Pierre 2000. *Férfiuralom*. Ford. N. Kiss Zsuzsa. Budapest, Napvilág.
- Bradiotti, Rósi 2007. Feminista filozófiák. In *uő Egy nomád térképei*. Ford. Vándor Judit és Agárdi Izabella. Budapest, Balassi. 7–33.
- Butler, Judith 1987. *Subjects of Desire. Hegelian reflection in Twentieth Century France*. New York, Columbia.
- Butler, Judith 1989. Sexual Ideology and Phenomenological Description. A Feminist Critique of Merleau-Ponty's Phenomenology of Perception. In Jeffner Allen – Iris Marion Young (szerk.): *The Thinking Muse – Feminism and Modern French Philosophy*. Bloomington, Indiana University Press. 85–101.
- Butler, Judith 1990/2006. *A problémás nem. Feminizmus és az identitás felforgatása*. Ford. Berán Eszter és Vándor Judit. Budapest, Balassi.
- Butler, Judith 1993/2005. *Jelentős testek. A „szexus” diszkurzív korlátairól*. Ford. Barát Erzsébet és Sándor Bea. Budapest, Új Mandátum.
- Butler, Judith 1997. Esetleges alapok. A feminizmus és a „posztmodern” kérdés. *Thalassa*. 1997. Újabb megjelenés in Csabai–Erős 1997. 256–275.
- Butler, Judith – Scott, Joan Wallach 1992. *Feminists Theorize the Political*. New York, Routledge.
- Csabai Márta – Erős Ferenc (szerk.) 1997. *Freud titokzatos tárgya. Pszichoanalízis és női szexualitás*. Budapest, Új Mandátum.
- Dinnerstein, Dorothy 1976. *The Mermaid and the Minotaur: Sexual Arrangement and Human Malaise*. New York, Harper.
- Eichenbaum, Louise – Orbach, Susie 1983. *Understanding Women. A Psychoanalytic Approach*. New York, Basic Books.
- Embre, Lester – Fisher, Linda (szerk.) 2001. *Feminist Phenomenology*. Dordrecht, Kluwer.
- Ferry, Luc 1992/1994. *Új rend: Az ökológia*. Ford. V. Tóth László. Budapest, Európa.
- Fraser, Mariam – Monica Greco 2005. *The Body. A Reader*. London, Routledge.
- Gay, Peter 1984. *The Bourgeois Experience: Victoria to Freud*. 1. köt. *Education of the Senses*. New York, Oxford University Press.
- Gaitens, Moira 1996. *Imaginary Bodies. Ethics, Power and Corporeality*. London, Routledge.
- Grosz, Elizabeth 1994. *Volatile Bodies. Toward a Corporeal Feminism*. Bloomington, Indiana University Press.
- Giuliani, Regula 2001. Das leibliche Selbst. Grenzen der Konstruktion des Geschlechts. In Waniek 2001. 205–219.

- Heinämaa, Sara 2003. *Toward a Phenomenology of Sexual Difference. Husserl, Merleau-Ponty, Beauvoir*. Lanham Boulder, Rowman.
- Hekman, Susan 1998. Material Bodies. In Welton 1998. 61–71.
- Hell Judit 2006. *Feminista filozófia*. Budapest, Áron.
- Irigaray, Luce 1974/1985. *Speculum of the other Woman*. Ford. Gillian C. Gill. Ithaca, Cornell.
- Irigaray, Luce 1993. *An Ethics of Sexual Difference*. Ford. Carolyn Burke – Gillian C. Gill. London, Athlone.
- Joó Mária 2003. Simone de Beauvoir és A második nem. *Pro Philosophia Füzetek*. 35. 35–57.
- Joó Mária 2004. A nő mint igazság Nietzsche-nél. *Pro Philosophia füzetek*. 40. 63–87.
- Joó Mária 2005. Simone de Beauvoir és a posztoszocialista helyzet. In Palasik–Sipos (szerk.) *Házastárs? Munkatárs? Vetélytárs? A női szerepek változása a 20. században*. Budapest, Napvilág, 39–55.
- Kruks, Sonia 2001. *Retrieving Experience. Subjectivity and recognition in feminist Politics*. Ithaca, Cornell.
- Moi, Toril 1999. What is a Woman? Sex, Gender and the Body in Feminist Theory. In uő *What is a Woman? and Other Essays*. Oxford, Oxford University Press. 3–121.
- Morgan, Katherine Pauly 1998. Sculpted bodies. Women and the knife: Cosmetic surgery and the colonization of women's bodies. In Welton 1998. 325–347.
- Mulvey, Laura 2000. *A vizuális élvezet és az elbeszélő film*. Ford. Juhász Veronika. *Metropolis, feminista különszám*. 4. 12–23.
- Nagl, Herta 2006. *Feminista filozófia*. Ford. Hell Judit. Budapest, Áron.
- Nussbaum, Martha 1999. The Professor of Parody. The Hip Defeatism of Judith Butler. *The New Republic*. Február 22. 37–45.
- Olkowski, Dorothea (szerk.) 2006. *Feminist Interpretations of Merleau-Ponty*. Bloomington, Indiana University Press.
- Pieper, Annemarie 2004. *Van-e feminista etika?* Ford. Hell Judit. Budapest, Áron.
- Preston, Beth 1996. Merleau-Ponty and Feminine Embodied Existence. *Man and World*. 29. 167–186.
- Price, Janet – Margrit Shildrick (szerk.) 1999. *Feminist Theory and the Body*. Edinburgh University Press.
- Séllei Nóra (szerk. és ford.) 2006. *A feminizmus találkozásai a posztmodernnel*. Debrecen, Csokonai.
- Salih, Sara 2003. *Butler*. London, Routledge.
- Suleiman, Susan Rubin (szerk.) 1985. *The Female Body in Western Culture*. Cambridge, Harvard University Press.
- Scott, Joan Wallach 2001. *Társadalmi nem – a történeti elemzés hasznos kategóriája*. In Joan Wallach Scott (szerk.) *Van-e a nőknek történelmük?* Ford. Greskovits Endre. Budapest, Balassi. 126–161.
- Young, Iris Marion 1990. *Throwing Like a Girl and Other Essays in Feminist Philosophy and Female Body Experience*. Bloomington, Indiana University Press.
- Young, Iris Marion 1990b. *Five Faces of Oppression in Young: Justice and the Politics of Difference*. Princeton, Princeton University Press.
- Young, Iris Marion 2005. *On Female Body Experience. 'Throwing like a Girl' and Other Essays*. Oxford, Oxford University Press.
- Vermes Katalin 2006. *A test éthosza*. Budapest, L'Harmattan.
- Waldenfels, Bernhard 2005. *A normalizálás határai*. Ford. Csátár Péter és Kukla Krisztián. Budapest, Gond-Cura.
- Wanick, Eva – Silvia Stoller 2001. *Verhandlungen des Geschlechts. Zur Konstruktivismusdebatte in der Gender-Theorie*. Wien, Turia und Kant Verlag.
- Welton, Donn (szerk.) 1998. *Body and Flesh. A Philosophical Reader*. Oxford, Blackwell.
- Wolf, Naomi 2000. *A szépség kultusza*. Ford. Follárdt Natália. Debrecen, Csokonai.

## A spinozai test mint probléma

Napjaink Spinoza-kutatásában hangsúlyosan jelen van egy olyan áramlat, amely nem annyira filozófiatörténeti érdeklődéssel közelít Spinozához és a rendszerével kapcsolatos kérdésekhez, mint inkább a spinozai módszer lehetséges alkalmazása és kiaknázása szándékával. Pierre-François Moreau egyenesen „forradalomról” beszél ennek kapcsán, mely valójában már Alexandre Matheron, Gilles Deleuze, Étienne Balibar és Pierre Macherey munkásságával megindult, de az ezredfordulóra vált olyan önálló tendenciává, amely spinozista szemléletű kutatási programokat követelt ki a szociológia, politológia, ökonómia területein és határterületein.<sup>1</sup> Ha, mint gyakran előfordul, ezek a kutatások a spinozai test, illetve individuum fogalmaival elemzik e területek releváns jelenségeit – a különféle társadalmi csoportok individuációját, a hatalmi, termelési, fogyasztói viszonyok önszerveződését, a pénz- és tőkefolyamatokban tetten érhető *conatus* stb. –, úgy nem tűnik érdektelennek vizsgálat alá vonni magában a spinozai testfogalomban rejlő potenciált a rendelkezésünkre álló spinozai szövegek fényében. Spinoza híres-hírhedt kérdése – vajon mire képes a test? –, amióta Deleuze reflektorfénybe állította, *magának* a spinozai kérdésnek számít a test kapcsán. Mi az alábbiakban elsősorban az *Etika* szövegére támaszkodva, még a test képességeire irányuló kérdés előtt azt próbáljuk meg problematizálni, hogy vajon mi is a test Spinozánál.

### I. TEST: EGYEDI DOLOG ÉS INDIVIDUUM

Önmagában erre a kérdésre persze nem nehéz felelni, hiszen a test, mint tudjuk, a kiterjedésnek a modusza, amint ezt az *Etika* II. részének 1. definíciója ki is mondja. A spinozai lexika azonban gazdagabb ennél, s a modusz fogalom mellett használja az egyedi dolog, illetve az individuum fogalmait is, amelyekről

A tanulmány a 67798 számú OTKA-pályázat támogatásával készült.

<sup>1</sup> Lásd különösen Citton–Lordon 2008.

rádadásul külön definíciókat ad az *Etika* ugyanezen részében. Hogy mivel jár a test modusz volta, nagyjából ismert: a test a végtelen terjedelmű természetnek egy véges módosulása, a végtelen óceánnak egy vízfodra.<sup>2</sup> Ám hogy mivel jár az, hogy a test egyedi dolog is, illetve individuum is, már kevésbé ismert. Ezért először ennek fogunk utánajárni.

Spinoza az *Etika* II. részének 7. definíciójában határozza meg az egyedi dolog (*res singularis*) fogalmát:

Egyedi dolgokon olyan dolgokat értek, amelyek létezése véges és meghatározott. Mert ha több individuum egyetlen cselekvésben olyképpen hat össze, hogy valamennyien egyszerre egyetlen okozat okát alkotják, ennyiben ezeket összességükben egy egyedi dologként tekintem.

Ez a definíció három észrevételre adhat okot. Egyrészt a definíció láthatólag egy nominális definiensből és egy genetikus definiensből tevődik össze. Az első mondat a „véges és meghatározott létezés” jellemzőivel fogalmilag azonosítja a definiendumot, azaz nominális definíciót ad,<sup>3</sup> míg a második mondat azt határozza meg, hogy konkrétan hogyan, milyen körülmények között jön létre a definiendum, azaz genetikus definíciót ad. Második észrevételünk az, hogy a definícióban az egyedi dolgot nem úgy kapjuk meg, mint ami egy benne rejlő szubsztanciális egység jóvoltából válna egyedi dologgá, még csak nem is a számszerű egysége, részeinek száma vagy a térbeli lokalitása teszi egyedi dologgá, hanem a *cselekvés* egysége: *una actio*. A definíció sodorvonalában az egység eredetileg a cselekvést jellemzi („egyetlen cselekvésben”), majd áttevődik a cselekvés által kiváltott okozat egységére („egyetlen okozat”), majd innen a dolog egységére („egy egyedi dolog”). Azért lehetséges tehát egy egyedi dologról beszélni, mert több individuum egyetlen cselekvésben összehat: *in una actione concurrant*. Így ha az utcán egy lerobbant autó eltolásánál alkalmi segítséget nyújt néhány járókelő, akkor ők egyetlen egyedi dologgá állnak össze – nevezhetjük ezt „mentőcsapatnak” –, amennyiben egyetlen cselekvésben, az autó mozgatásában összehatnak. Sarkosabban úgy is fogalmazhatnánk, hogy a természetben ekkor *keletkezik* egy egyedi dolog, mely a cselekvés végeztével, miután a járókelők folytatják a maguk útját, *megszűnik létezni* – teljes összhangban a definíció ama kitételével, hogy a dolog létezése véges és meghatározott. Végül a harmadik észrevételünk az, hogy az egyedi dolog definíciója a cselekvésen túl támaszkodik az *individuum* fogalmára is. Nézzük meg ezt közelebbről.

Az *individuum* fogalma szórványosan felbukkan ugyan korábban az *Etika* I. részében, de igazán csak a későbbiekben kap hangsúlyos szerepet, mi több, önálló

<sup>2</sup> A vízről mint a szubsztancia–modusz viszony metaforájáról lásd az *Etika* I. 15. tétel megjegyzését.

<sup>3</sup> Amit voltaképpen már az *Etika* I. könyv 28. tétele is kimond: „Minden egyedi dolog, vagyis minden dolog, amelynek létezése véges és határolt...”

definíciót, az *Étika* II. részének fizikai betétjében, a 3. segédétel következtetett tétele után. Ez a definíció az individuum fogalmát kifejezetten a fizikai testek síkján, azaz kizárólag a kiterjedés attribútumára leszűkítve határozza meg:

Mikor több egyforma vagy különböző nagyságú testet más testek olyképp szorítanak össze, hogy egymást érintik, vagy ha ugyanolyan vagy különböző sebességgel mozognak olyképpen, hogy ezt a mozgásukat bizonyos meghatározott arány (*ratio*) szerint közlik egymással, akkor azt mondjuk, hogy ezek a testek egyesültek egymással, s mind együttesen *egyetlen testet vagy individuumot* alkotnak, amely a testeknek ezen egyesülése által különbözik a többitől.

Az individuum definíciója immár ízig-vérig genetikus definíció, tehát az individuum, a testi individuum keletkezését tárja elénk: egy individuum keletkezhet egymást érintő, egymáshoz tapadó testekből, de keletkezhet úgy is, hogy e testek nem összetapadnak, hanem mozognak egymáshoz képest. Ez utóbbi esetben a meghatározó feltétel az, hogy (a) e mozgások nem függetlenek egymástól, hanem „kommunikálnak” (*communicent*), vagyis egy bizonyos mozgásközlési rendszerbe szerveződnek, melyben mozgásokat adnak át egymásnak; (b) továbbá ez a mozgásközlési rendszer egy meghatározott aránynak vagy viszonynak – *ratió*nak – engedelmeskedik (*certa quadam ratione communicent*). Az első eset triviális: voltaképpen már Descartes is erre vezette vissza a testeket, még ha ő nem is kifejezetten a test részeinek összetapadására helyezi a hangsúlyt, hanem arra a tényre, hogy a részei együtt mozognak, szimultán mozdulnak el.<sup>4</sup> A második esetben, amellyel Spinoza már túllép Descartes-on, aki a részek egymáshoz képest tekintett mozgását soha nem fogja konstitutív tényezőnek tartani, azt látjuk, hogy akár az (a), akár a (b) feltétel elmaradása esetén *nem keletkezik individuum*: a levegőben kavargó porszemek vagy a vízben úszkáló halak, pusztán azáltal, hogy egymáshoz képest mozgásban vannak, még nem fognak egyetlen individuummá összeállni (sem [a], sem [b] nem teljesül). De még akkor sem, ha a porszemek egymással összeütközve mozgást közölnek egymással, vagy ha a halak a véletlenszerű mozgásaikat egymáshoz képest úgy szervezik meg, hogy kitérjenek egymás elől, és ne ütközzenek össze (az [a] teljesül, [b] nem). Individuum csak akkor keletkezik, ha e mozgásközlési rendszerekben a mozgások közt egy meghatározott arányosság vagy szabályszerűség lép fel: ha például a porszemek örvénylő kavargásba kezdenek, vagy ha a halak mozgása összehangolódik, például egy vadászó vagy menekülő formáció szerint.

<sup>4</sup> „Testen, vagyis anyagrészen értem azt az összességet, ami a mozgás során egyszerre mozdul el, függetlenül attól, hogy az összesség részei eközben egyéb mozgásokat végeznek vagy sem.” Descartes 1647/1997 II. 25. 170.

Egymás mellé állítottuk az egyedi dolog és az individuum definícióját. Kérdésünk az, hogy ha most eltekintünk attól, hogy egy testre hivatkozhatunk egyedi dologként is és individuumként is, önmagában e két meghatározás között milyen kapcsolatot tudunk teremteni. Végezzünk ehhez némi diszkussziót.

Egyfelől érvelhetünk amellett, hogy az individuum definíciója konkretizálja az egyedi dolog definíciójának a tartalmát: amit az első egységes cselekvésként jelölt meg (*in una actione concurrant*), az a másodikban egy bizonyos meghatározott arány szerint zajló mozgásközlésként nyer konkrétabb tartalmat (*motus suos invicem certa quadam ratione communicent*). Akár még odáig is elmehetnénk, hogy azt mondjuk, az első definícióban szereplő *cselekvésnek* a második definíció *mozgásközlése*, a cselekvés *egységének* pedig a másodikban az a bizonyos *arány* felel meg. Eszerint az individuum fogalma konkrétabb és gazdagabb körülírását adja az egyedi dolog fogalmának, ám voltaképpen individuum és egyedi dolog ugyanaz: e fogalmak szinonimák. Egymásra következésük rendje az *Etika* szövegében jól megfelel annak a spinozai tendenciának, amely a megismerés általánosabb formáitól az egyre konkrétabb és specifikusabb formák felé visz bennünket, más szóval a megismerés második nemétől a harmadik nem, a konkrétumok intuitív ismerete felé vezet.

Ugyanakkor az individuum és egyedi dolog ilyenén azonosítása mégsem teljesen magától értetődő, és akár még az a gyanú is felmerülhet, hogy e definíciókban a sokaságok két eltérő organizációjával van dolgunk. Hiszen mit is állítunk akkor, ha az individuumról azt mondjuk, hogy egyedi dolog? Ha egy individuum egyedi dolog is, azaz kielégíti az egyedi dolog általánosabb definícióját, akkor ez azt jelenti, hogy olyan individuumokból kell állnia, amelyek egyetlen cselekvésben összehatva egyetlen okozatot hoznak létre, legyen az bármi. Más szóval, nem csupán annak kell teljesülnie, amit az individuum definíciója kimond, hogy az őt alkotó individuumok között arányos mozgásközlés zajlik, hanem ezenfelül az arányos mozgásközlésnek ki kell váltania egy okozatot, *egyetlen* okozatot. Spinoza például az emberi testet individuumnak tekinti: a sejtek, a vér, a nyirok, a szív, az agy stb. mint megannyi individuum egymáshoz képest arányosan megszerveződő milliárdnyi mozgásaikkal létrehoznak egy nagyobb, komplexebb individuumot, a testünket. De vajon kiváltanak-e ezek a mozgások együtt egyetlen okozatot? És mi lenne ez az okozat? A test sokféle okozat oka lehet, de ha mozdulatlanul alszik, akkor is megmarad individuumnak, pedig ekkor semmilyen okozat kiváltásában nem tűnik cselekvő ágensnek.<sup>5</sup> Így ha elkötelezzük magunkat amellett, hogy részindividuumainak mozgásai okozatot

<sup>5</sup> Eltekintve persze attól, hogy lélegzésével mozgatja a környező levegőt, testmelegével hat a környezetére stb. Itt jegyezzük meg, hogy lehetséges volna az individuumot *kölcsönhatásként*, illetve merő *viszonylagosságként* elemezni, mely elemzésben az individuum feloldódna abban a karakterisztikumában, hogy csak a környezetéhez, illetve a természet egészéhez viszonyítva, e viszonyok sokaságában az, ami. Ez az elemzés azonban egy másik tanulmányt követelne.



váltak ki, úgy ezt az okozatot nem kereshetjük rajta kívül, a testen túl. Csak azt felelhetjük tehát, hogy maga az emberi test léte ez az okozat. Az *Etika* III. részének fogalmait segítségül hívva, a részindividuumok *részokai* vagy *inadekvát okai* egy olyan okozatnak, ami – a szó szoros értelmében – testet ölt. Mit jelent ez? Ha a korábbi példánkban a lerobbant autó *elmozdulása* egy okozott esemény volt a természetben, most az emberi test *aktuális létezését* ismerjük el eseményként. Ott az esemény egy kész testtel kapcsolatos esemény volt, itt maga az esemény „szüli meg” a testet. Végző soron tehát ha az individuumokat egyedi dolgoknak nevezzük, ezzel eseményként tekintünk rájuk: eseményként, amit sok-sok test együttes cselekvése vált ki, nevezzük így, egy inkorporáló kauzalitással, ami borotvaélen táncol a megszűnés és az átalakulás között. Megszűnik ugyanis, ha e sok-sok test önállóan, s nem összehangoltan kezd cselekedni. És átalakul, ha ugyanez a testsokaság együttesen és egyszerre, de *máshogyan* kezd el cselekedni.

Akárhogy is legyen, az individuum és az egyedi dolog azonosításával nyertünk egy új perspektívát a testről. Egy olyan perspektívát, amely ugyanakkor implicit marad a spinozai szövegekben, mivel mindig rávetül a *causa seu ratio* explicit perspektívája, amely szerint „minden dologra nézve meg kell tudnunk jelölni az *okot* vagy *alapot* (*causa seu ratio*), hogy miért létezik, vagy miért nem létezik” (*Etika* I. 11. tétel, 2. biz.). A test létének *ratio*ja ezen explicit perspektívában az individuum definíciójában kikötött *ratio* lesz, amely szerint a test részei kölesönösen mozgást adnak át egymásnak: a mozgásközlés bizonyos meghatározott aránya. E meghatározott arány miatt létezhet egyáltalán *ilyen* test a természetben, *ilyen* tulajdonságokkal és *ilyen* képességekkel. Ez az arány determinálja a test létét. Ugyanakkor maga az arány is olyan, hogy egy bizonyos anyagdarab (pontosabban testsokaság) fel tudja venni azt, hiszen az arányok „az egyetemes testi természet rendjébe” illeszkednek: felvehetők. Spinoza szerint persze az emberi elme is kiötölhet mindenféle tetszetős arányszüleményeket, ám előfordulhat, hogy ezek semmilyen test okául vagy alapjául nem szolgálhatnak, semmilyen testsokaság nem képes felvenni őket, s így természetellenes kimératestnek bizonyulnak. Ahogyan vannak gondolati képtelenségek (szögletes kör), úgy vannak testi képtelenségek is (griffmadár, szárnyas ló, beszélő fa). Ha a *causa seu ratio* perspektívájában soha nem fog gondot okozni a valóságos testek megkülönböztetése a kimératestektől, az azért van, mert e perspektívában a test *ratio*ja és az egyetemes testi természet rendje cirkulárisan megalapozza egymást, egymás függvényében érvényesek. Ezzel szemben ha az inkorporáló kauzalitást explicitté, önálló szemponttá tennénk – ahogyan nem az –, azt kellene mondanunk, hogy egy fokkal kevesebbet tudunk: e perspektívából csak annyit látunk, hogy sok-sok test együttesen cselekszik, ám nem feltétlenül pillantjuk meg a cselekvés *ratio*ját, a mozgásközlés meghatározott arányát, így azt sem tudjuk megmondani, hogy *miféle* test keletkezett, illetve *egyáltalán* test keletkezett-e. Kilépnünk a cirkuláris megalapozásból, így nem tudjuk *elcse* megkülönböztetni a

kiméreteket a valódi testtől. Az eseményt, sok-sok test összehatását, meg kell fejtenünk. Előre sejtethető így az is, hogy a *causa seu ratio*ból és az inkorporáló kauzalitásból komplementer problematizációs csapások fognak következni: az elsónél egy létező testből indulunk ki és keressük azt a karakterisztikus arányt, amely létének oka és alapja; míg a másodiknál egy mozgásközlési rendszerből, azaz sok-sok egyszerre cselekvő testből indulunk ki, és keressük a rendszer karakterisztikus arányát, vagyis azt a nagyobb, komplexebb testet, amit e rendszer konstituál. A második út annyival rázósabb az elsónél, hogy míg az első esetben tudjuk, hogy léteznie kell valamilyen aránynak (hiszen a természet rendje biztosít efelől), addig a második esetben kérdéses, hogy *megfelel-e* test a kiinduló aránynak. Ezekre a kérdésekre a későbbiekben fogunk bővebben kitérni, most csak jó előre jelezzük őket.

Térjünk vissza az egyedi dolog és az individuum kapcsolatára, és nézzük meg a fordított irányú azonosítást, vagyis azt, hogy mivel jár, ha most az egyedi dolgot azonosítjuk individuumként. Ezzel ugyancsak zavaros vizekre evezünk. Nézzük ugyanis a korábbi autós példánkat: azt mondtuk, a lerobbant autót eltoló individuumok egységes cselekvésük jóvoltából egy egyedi dolgot alkotnak, egy mentőcsapatot. Ám ha a mentőcsapat individuum is, azaz kielégíti az individuum definícióját, akkor azt kell mondanunk, hogy az autót eltoló individuumok *között* mozgásközlés zajlik, hiszen a definíció ezt kötötte ki. Ezt azonban nemigen látjuk, hiszen az együtt ható individuumok nem egymással közölnek mozgást, hanem az autóval, amit mozgatnak. Ennek a definíciós feszültségnek az okára könnyű rámutatnunk: minden bizonnyal abban kell látnunk, hogy Spinoza az individuum meghatározásakor mindenekelőtt az emberi testet, vagyis egy komplex élő szervezetet tartott szem előtt, melynek karakterizálásánál kritikai hévvel utasította el a descartes-i „nyugvó tömeget” és a vele kapcsolatos szempontrendszereket.<sup>6</sup> E kritikai elutasítás következménye az, hogy „a részek szimultán elmozdulása”, mely a descartes-i testnek lényegi szempontja, Spinozánál csak a legprimitívebb testekre marad érvényes, tudniillik azokra a szilárd testekre, amelyekben a részek egymáshoz tapadnak, míg a komplexebb testek esetében ez a szempont értelmezhetetlenné, illetve valamiféle szélsőséges háttáresetté válik. Természetesen ettől még kísérletezhetünk mindenféle áthidaló interpretációval, ám látnunk kell, hogy egy ilyen feszültség megléte a spinozai rendszerben nem tekinthető szimpla negatívumnak, aminek a nyomait el kéne takarítanunk, hiszen a spinozai gondolkodás egy pozitív karakterisztikuma csilán meg benne, az, ahogyan Descartes-nak búcsút intve saját problémát állít középpontba, az individuáció és az organizáció problémáját.

<sup>6</sup> Vö. Spinoza 81. levelével Tschirnhaushoz: „A kiterjedésből, ha úgy fogjuk fel, mint Descartes, tudniillik mint nyugvó tömeget, a testek létezésének bizonyítása nemcsak nehéz, mint Ön mondja, hanem teljességgel lehetetlen.”

Ha áthidaló értelmezésekkel kísérletezünk, úgy tűnik, több lehetőség közül is választhatunk. Egyrészt értelmezhetjük az individuum definícióját úgy, hogy beleférjen egy szimultán elmozdulást produkáló mozgásrendszer is. Hiszen a *motus suos invicem certa quadam ratione communicent* formula eléggé tágan értelmezve annyit mond, hogy a részek a mozgásaikat közösen alakítják ki, közös nevezőre hozzák, egyesítik – a *communicentet* mindenesetre olvashatjuk így. A közösen kialakított mozgásba pedig már nyilvánvalóan belefér a részek szimultán elmozdulása is: az autót eltoló individuumok a mozgásaikat meghatározott módon közösen alakítják ki, anélkül hogy ezzel azt kellene állítanunk, hogy közöttük mozgásközlés zajlik. Másik lehetséges megoldás az, hogy kitágítjuk az egyedi dolog definícióját úgy, hogy az egyetlen okozat okát alkotó individuumokat *plusz magát ezt az okozatot* is tekinthessük összességében egy egyedi dolognak. Ez utóbbi esetben a *mentőcsapatot plusz az autót* tekintenénk egy egyedi dolognak, s így az individuum definíciójának kritériumára áttérve már nem azt kellene állítanunk, hogy a mentőcsapat tagjai között zajlik mozgásközlés, hanem azt, hogy a mentőcsapat és az autó között, így a mentőcsapat és az autó egy individuumot alkot. Még eklatánsabbá válik ez a példa, ha a mentőcsapatot kidobóemberekkel, az autót pedig egy kocsmatöltelékkel helyettesítjük, akit a kidobóemberek megpróbálnak kituszkolni, az pedig ellenáll: ez esetben az ok és az okozat a szemünk láttára individualizálódik egy „verekedő társasággá”, azaz egyetlen, jól meghatározott mozgásközlési rendszerbe szerveződő individuummá. Általánosságban minden ilyen összetapadó mozgó-mozgató konstellációt tekinthetünk egy individuumnak. Végül harmadik lehetőségként mondhatnánk azt, hogy a komplexebb individuumok esetében minden szimultán elmozdulás, amely egyszerre cselekvés és összehatás, feltételez transzverzális mozgásközléseket is, melyekre hivatkozva az összeható individuumokat immár teljes joggal definiálhatjuk egyetlen individuumként. Hiszen a lerobbant autót eltoló individuumok nem cselekedhetnének egyszerre, ha valaki nem jelezne közülük, hogy „Háromra toljuk meg!”, amint a halak sem úszhatnának egy rajban, ha nem figyelnék egymást, vagyis ha nem lenne közöttük valamilyen auditív, vizuális vagy bármilyen egyéb jelző kommunikáció, amit Spinoza, ne felejtjük el, *testi* kommunikációnak, vagyis testek közti mozgásközlésnek tekint. Ennek a szimultán elmozdulást átlósan keresztező mozgásközlésnek a jelentőségét mutatja az is, hogy milyen irtatlanul nehéz egyébként nagy komplexitású rendszerekben szimultán elmozdulást előidézni (lásd sztrájk, tüntetés, polgári engedetlenség stb.).

Áttekintettük az egyedi dologgal és az individuummal kapcsolatos szempontokat. A tapasztalt különbségek ellenére az is kiderült, hogy a test, már ha nem egy tökéletesen szilárd és merev testről van szó, a spinozai felfogás szerint mindig mozgást feltételez, mozgás révén válik testté. A továbbiakban ezt fogjuk alaposabban körüljárni.

## II. TEST MINT MOZGÁS ÉS MINT MOZGÁSOK ARÁNYA

A mozgásközlési rendszerről, amely az individuumokat definiálja, Spinoza változatos módon beszél, és még ennél is változatosabb ezeknek a szöveghelyeknek a magyar fordítása, amiben nem annyira az átültetés megoldatlanságát kell látnunk, mint inkább annak a tartalomnak a gazdagságát, ami a latin formulákba is csak nehezen sűrítendő bele. Az individuum definíciójában a *motus suos invicem certa quadam ratione communicent* formulát fentebb a „mozgásukat bizonyos meghatározott arány szerint közlik egymással” megfogalmazással adtuk vissza, eltérve az *Etika* Szemere Samu fordította, Boros Gábor által a közelmúltban alaposan átdolgozott új magyar szövegében szereplő „mozgásukat bizonyos meghatározott módon közösen alakítják ki” formulától. Tettük ezt azért, hogy a *ratio* jelenlétét hangsúlyozzuk, és láthatóvá tegyük a kapcsolatát a fizikai betét azon szöveghelyeivel, ahol Spinoza ugyancsak mozgások arányáról, viszonyáról beszél. Ugyanakkor azt is látjuk már, hogy ennek a fordításnak megvan az az előnye, hogy az egyedi dolog és az individuum definíciói közti feszültséget feloldja. Mindazonáltal ez a megoldás nem meríti ki teljesen a latin szöveg értelmét, amit az mutat a legjobban, hogy az *Etika* II. 24. tételének bizonyításában, ahol a latin formula lényegében szó szerint megismétlődik, az új magyar fordítás ezúttal így hangzik: „a mozgásukat bizonyos meghatározott szabály szerint közösen alakítják ki.” Vagyis most a *ratio*t szabályként, a mozgásátadás szabályaként értjük meg. Végül az *Etika* IV. 39. tételének bizonyításában a latin szöveg lényegében megint kimondja a kiinduló formulánkat, és az idézett fordítás ezúttal azt a verziót választja, amit mi is: a részek „meghatározott arány szerint közlik egymással mozgásaikat”. Már itt leszögezzük, mindegyik megoldásnak megvan a maga előnye, és a *ratio* különböző jelentéseit legjobb együttesen szem előtt tartani.

Nézzünk széjjel ezek után ismét a spinozai szövegben. A II. rész 5. segédtételeiben Spinoza azt mondja, hogy az individuum akkor őrzi meg a természetét, ha az alkotórészei „között a kölcsönös mozgás és nyugalom aránya ugyanaz marad”, *omnes eandem, ut antea, ad invicem motus, et quietis rationem servant*. Itt is felbukkan tehát a *ratio*, de ezúttal már maga a latin formula is változik, amennyiben nem a mozgásközlés, hanem a mozgás–nyugalom *ratio*járól szól, miközben világos, hogy a mozgásközlésnek és a mozgás–nyugalomnak valamiképpen szinonimáknak kell lenniük, hiszen a segédétel az individuum megmaradásáról szól, azaz az individuum részei mozgásaikat továbbra is ugyanazon meghatározott arány szerint közlik egymással, amit az individuum definíciója megszab. Ezt támasztja alá az *Etika* IV. 39. tételének már említett bizonyítása is, ahol Spinoza együtt szerepelteti e két kifejezést – tudniillik az emberi test „részei meghatározott arány szerint közlik egymással mozgásaikat”; illetve „a mozgás és nyugalom aránya megmarad az emberi test részeinek egymáshoz való viszonyában” –, és kifejezetten szinonimaként kezeli őket, lévén, hogy mindkettőt azt a feltételt fogalmazza meg, amelynek teljesülése esetén az illető test fennmarad, természete megőrződik.

Itt érdemes utalnunk arra is, hogy a „mozgás–nyugalom aránya” formula időrendben megelőzi az „arányos mozgásközlés” formuláját, ugyanis Spinoza korábban, a *Rövid tanulmányban* a testek létét, fennmaradását, majd változását és pusztulását már a mozgás–nyugalom arányára hivatkozva határozta meg, a másik formula említése nélkül:

Minden különös dolog, amely létezéshez jut, mozgás és nyugalom által lesz ilyenné, s így van a szubsztanciális kiterjedés valamennyi modusza, amelyet *testnek* nevezünk. A testek különbsége egyedül a mozgásnak és nyugalomnak mindig más aránya által keletkezik. Innen van, hogy ez a valami *ilyen* és nem amolyan, *ez* és nem más. [...] Azonban a mozgásnak és nyugalomnak más arányait mutatta testünk mint újszülött gyermek, másokat a későbbi életkorban, s ismét más arányokat mutat majd, ha meghaltunk. [...] Ha mármost ilyen testnek ez az aránya – például 1:3 –, s ezt megtartja, akkor a lélek és a test olyan lesz, mint most a miénk, folytonos változásnak alávetve ugyan, de nem olyan nagy változásnak, hogy meghaladná az 1:3 határait. [...] Ha azonban más testek oly erővel hatnak a miénkre, hogy a mozgásának 1:3 aránya nem maradhat meg, akkor ez a halál... (*Rövid tanulmány* II. rész, bevezetés, 7, 8, 10, 12, 14. jegyzet).

E sorokból az is látszik, hogy mennyire konkrétan, milyen számszerű konkrétsággal gondolkodott Spinoza a *Rövid tanulmányban* az arányról. A mozgásnak és a nyugalomnak az 1:3 aránya persze csak illusztráció, ami mögé még annyi szemléletes tartalmat is nehéz gondolnunk, mint amivel a késő reneszánsz antropometrikai arányelméleteiben találkozunk a vitruviusi tanokat továbbfejlesztő Albertinél, Leonardónál vagy Dürernél – ám azt azért megállapíthatjuk, hogy az arányossági diskurzust Spinoza többnek tartotta pusztá metaforánál, és a kvantitatív jegyek felmutatásával egy olyan filozófiai fogalmat alkotott, amely természettudományi tekintéllyel és relevanciával is bír, azaz a természet egy lényegi mozzanatára mutat rá.<sup>7</sup>

Térjünk vissza az *Etikához*. A II. részben található fizikai betét első két axiómájából nemcsak az világos, hogy mozgás és nyugalom két állapot, amit egy test felvehet, hanem az is, hogy a testek képesek gyorsabban és lassabban mozogni, felgyorsíthatók és lelassíthatók. Joggal biggyeszti ezért Spinoza a fizikai betét néhány segédételében a *motus et quies* mellé a *celeritas et tarditas* fogalompart, amiből azt a következtetést is levonhatjuk, hogy az arány kívánalma a mozgás–nyugalmon túl a gyorsaságra és lassúságra, felgyorsításra és lelassításra is kiterjedhet. Valóban, a 7. segédételt követő megjegyzés, melyben Spinoza összefoglalja és lezárja azt, amit az individuumok fizikájáról tudnunk kell, a természetet nö-

<sup>7</sup> A *Rövid tanulmány* és az *Etika* arányfogalmának összevetéséről lásd bővebben Zourabichvili 2002. 51–92.

vekvő komplexitású individuumok hierarchiájaként láttatja, ahol egy bizonyos szint után olyan individuumokkal találkozunk, melyeknek „minden egyes része saját természetének minden változása nélkül majd lassabban, majd gyorsabban mozoghat, s következőleg mozgását majd lassabban, majd gyorsabban közölheti a többivel”. Ezeknél a komplexebb individuumoknál az individuáló arány már nem korlátozódhat a pusztá sebességi viszonyokra, hanem a sebességváltozási viszonyokat is magában kell foglalnia.

Végül e megjegyzéssorozatot lezárandó, állapítsuk meg, hogy Spinoza használja a „mozgás–nyugalom törvényei” kifejezést is az *Étika* III. 2. tételének megjegyzésében, mint amely törvények a testeket a kiterjedés attribútumában meghatározzák, ilyen vagy olyan viselkedésre determinálják.<sup>8</sup> Ez a kifejezés nem csupán arra enged következtetni, hogy az egyes individuumokban a mozgás–nyugalom arányai (*ratio*) mind ugyanazon átfogó mozgás–nyugalmi törvényeknek (*leges*) engedelmessé válnak, mely utóbbiak természettörvényekként, mechanikai törvényekként funkcionálnak,<sup>9</sup> hanem azt is lehetővé teszi, hogy a kifejezés első tagját, a „mozgás–nyugalom” fogalmát egyszerűen azonosítsuk: a „mozgás–nyugalom” kifejezésen Spinoza azt érti, amit ma egyetlen szóval dinamikának neveznénk. Dinamikai törvények szabályozzák azokat a dinamikai arányokat vagy viszonyokat, amelyek mindegyikének egy-egy individuum felel meg a természetben.

Milyen tanulságot szűrhetünk le ezekből a megjegyzésekből? Egyrészt azt, hogy a *ratio* jelentése a „viszony”-tól, „arány”-tól a „szabály”-ig terjedő teljes skálát átfogja: a spinozai megfogalmazásokat joggal értelmezhetjük úgy is, hogy az egyes mozgások, külön-külön tekintve, meghatározott arányban vagy viszonyban állnak egymással, és úgy is, hogy a mozgások meghatározott szabályszerűség szerint kapcsolódnak össze. Másrészt mivel a szövegek, különösen a már idézett 7. segédteletet követő megjegyzés, mozgás és nyugalom mellett gyorsaságra és lassúságra is hivatkoznak, ezért a szóban forgó *ratio* (viszony, arány, szabály) hatályát a monoton és változatlan mozgások összevetésén túl a mozgás-változásokra – lassulásokra, lelassításokra és gyorsulásokra, felgyorsításokra – is kiterjeszthetjük. Vagyis olyan megállapításokat is tehetünk, hogy az egyes mozgások változása, külön-külön tekintve, meghatározott arányban vagy viszonyban áll egymással, illetve a mozgásközlési rendszerben tapasztalható (lokális vagy

<sup>8</sup> „Az elme elhatározása is, vágya is, s a test determinációja természet szerint együtt jár, helyesebben, egy s ugyanaz a dolog; ha ezt [...] a kiterjedés attribútumában tekintjük, s a mozgás és nyugalom törvényeiből (*ex legibus motus et quietis*) származtatjuk, determinációnak nevezzük.”

<sup>9</sup> Spinoza Oldenburghoz írt 13. levelében ír mechanikai törvényekről: „a testek minden változása mechanikai törvények szerint megy végbe.” Uthalhatunk még itt az *Étika* I. Függelékére, ahol Spinoza azokon a balgákon ironizál, akik az emberi test bámulatos felépítéséből „arra következtetnek majd, hogy a szervezet nem mechanikus úton jött létre, hanem isteni vagy természetfeletti művészet alkotta meg”. A spinozai álláspont szerint nyilvánvalóan a legbámulatosabb test is csak és kizárólag mechanikus úton jöhet létre.

globális) lassulások, illetve gyorsulások meghatározott szabályszerűségnek engedelmessé válnak. Bármelyik megfogalmazást válasszuk is, megállapíthatjuk, hogy az, ami egy individuumot individualizál és a maga individualitásában összetart, nem a részeiben keresendő, hanem abban, hogy a részei felveszik-e és megőrzik-e ezt a bizonyos dinamikai viszonyt vagy szabályszerűséget. Hiszen ha ugyanazok a részek egy egészen más viszony vagy szabály szerint kezdenek el mozogni, az individuum megszűnik létezni, és a helyén egy másik (vagy akár több másik) individuum keletkezik, melynek (vagy melyeknek) individualitása pontosan ebben az új viszonyban vagy szabályban fog állni.

Összefoglalásképpen nem lesz haszontalan felsorolni, hogy a mozgások *ra-tiójától*, dinamikai szabályától függően milyen típusú testekkel találkozunk a spinozai természetben:

1. Tekintsünk egy tömör testet, egy követ például. A kő részéről elmondhatjuk, hogy összetapadnak, de azt is mondhatjuk, hogy nyugalmi viszonyban vannak egymáshoz képest: mindkét alapon egyetlen testet alkotnak. A kő dinamikai szabálya a nyugalom, a részek kölcsönös nyugalma. Ha a részek tapadását, nyugalmi viszonyát valamilyen ok megszünteti (a követ egy nagyobb kő összezúzza, vagy tűz hatására a részei mozgásba jönnek és elválnak egymástól stb.), a kő átalakul.

2. A második típusba sorolhatjuk azokat a problémás individuumokat, amelyek már nem tömör testek, de az egyes részek nyugalmi viszonyban vannak egymáshoz képest. Ilyen egy széllökés, egy halraj, amely a vízben ide-oda cikázva folyamatosan együtt marad, és egy testet alkot, ilyennek tekinthető az autót mozgató mentőcsapat vagy egy egyszerre menetelő embertömeg. Itt már semmiképpen sem beszélhetünk összetapadásról, viszont korábbi elemzéseink fényében elmondhatjuk (a) vagy azt, hogy ennek a testnek a dinamikai szabálya is a nyugalom, (b) vagy azt, hogy a test dinamikai szabályát a nyugalom és a transzverzális mozgások kompozíciója alkotja.

3. Tekintsünk egy órát. Az óra egyes fogaskerekei meghatározott sebességgel forognak egymáshoz képest, illetve az óra minden egyes alkatrészéhez (ingához, vázhoz stb.) képest is, miközben szabályos és meghatározott módon mozgást adnak át egymásnak. Az egyes alkatrészek sebességkülönbségeit akár számszerűen is ki lehetne fejezni sok-sok aránypárral (hasonlóan a *Rövid tanulmány* példájához), vagy akár egyetlen, rendezett és valószínűleg igen hosszú aránypárformulával: az óratest dinamikai szabálya pontosan ez a rögzített aránypárformula lesz. Úgy is fogalmazhatnánk, hogy ez a formula definiálja az illető individuumot, megadja a természetét, és azt, hogy mi következik e természetből, mire képes e természet. Egy másik aránypárformula egy másik individuumot definiálna, amely lehet, hogy óráként működne, de az is lehet, hogy egészen furcsán és másként működne, vagy pedig egyáltalán nem működne. Vajon az égi órákat is tekinthetjük ilyen individuumnak? A Naprendszer, középen a Nappal, körülötte a más-más sebességgel mozgó bolygókkal, a körülöttük keringő holdak-

kal nyilvánvalóan értelmezhető egymáshoz képest meghatározott viszony vagy arány szerint mozgó testek rendszereként, vagyis a Naprendszer leírható egy aránypárformulával. De vajon individuum-e a Naprendszer? Azaz vajon mozgásközlési rendszer-e ez a rendszer? Mivel Spinoza csupán egy vázlatos és vállaltan töredékes fizikát vetett papírra az *Etika* lapjain, és kozmológiai elképzeléseit máshol sem fejtette ki, ezért ha egyáltalán beszélhetünk átfogó fizikai világmépről Spinozánál, ezt a világmépet csak aprólékos elemzésekből nyert szilánkokból állíthatjuk össze. A kérdést így függőben hagyjuk.

4. Tekintsük az emberi testet. Emlékezzünk vissza, hogy az *Etika* II. részének egész fizikai betéte azt a célt szolgálja Spinoza számára, hogy egy emberi test komplexitású individuumról számot tudjon adni, és ezen individuum működését párhuzamba tudja állítani a hozzá tartozó elme működésével. Az emberi test, benne a lüktető szívvel, a keringő vérrrel, a saját mozgással rendelkező és különféle változatos mozgásokat továbbadó szervek sokaságával immár egy olyan komplexebb individuumtípust alkot, amelynek a dinamikai szabályát nem tudjuk megadni az óráéhoz hasonló konstans aránypárformulával. Hiszen a szervek mozgása egymáshoz képest, valamint az egymásnak átadott mozgások nem egy konstans mentén szerveződnek meg, hanem maguk is változnak, gyorsulnak és lassulnak, gyorsítanak és lassítanak, ráadásul úgy, hogy e változások, illetve a változás szélső értékei szervről-szervre más-más szabályoknak engedelmessé válnak. Az emberi test dinamikai szabálya tehát nem fejezhető ki számszerű aránypárformulával: talán ennek felismerése miatt tűnik el az *Etikából* a *Rövid tanulmány* matematikai illusztrációja (az 1:3 arány), jóllehet az sem tűnik képtelenségnek, hogy Spinoza értelmesnek tekintett volna egy olyan formulát, amely nem magukat a változásokat számszerűsíti, hanem a változások szélső határait és a köztük lévő kapcsolatokat.<sup>10</sup> Az mindenesetre biztos, hogy akár számszerűsíthető, akár nem, a test részei meghatározott arány szerint közölnek egymással mozgásokat, vagyis a testnek van egy dinamikai szabálya, amely individualizálja őt, definiálja a természetét, és azt, hogy

<sup>10</sup> Vö. Spinoza Meyerhez írt 12. levelével, ahol Spinoza azon dolgok közt, amelyek „természete nyilvánvaló ellentmondás nélkül nem tűrheti meg a számot”, felsorolja a változást is: „valamennyi változás, amelyet egy [...] mozgó anyagnak el kell szenvednie, meghalad minden számot.” A *Celeritatis et tarditatis* tehát számmal meghatározhatatlan. Ugyanakkor Spinoza jóváhagyja, hogy a meghatározatlan természetek „maximumát és minimumát ismerjük”, továbbá a nagyobb–kisebb relációt alkalmazhatjuk rájuk. Uthalhatunk még itt Deleuze interpretációjára, aki egyik előadásában a mozgás–nyugalom viszonyát a testet felépítő végtelen sok rész sebességkülönbségeinek egy differenciálhányados formájában végessé és számszerűvé váló viszonyaként értelmezi (Deleuze 1981. 02. 17). Ehhez az intuitív meggyőző szemléltetéshez csak azt szükséges hozzátennünk, hogy a Meyer-level tanúsága szerint Spinoza alapvetően bizalmatlan volt a meghatározatlan természetek matematizálásával szemben – túl azon, hogy a differenciálkalkulussal nem is találkozhatott, hiszen annak kimunkálásán Newton a Meyer-level írása körül kezd el dolgozni, és az eredményeit is csak jóval ezután hozza nyilvánosságra, azt követően, hogy Leibniz 1684-ben megjelenteti a maga kalkulusát az *Acta Eruditorum* hasábjain. Spinoza ekkor már rég nincs az élők sorában.



mire képes e természet. Nem kétséges, hogy a felmutatható közös mozgásminták (vérkeringés, szívdobogás stb.) ellenére minden embernek más és más a dinamikai szabálya, és Spinozát – a megismerés harmadik nemét kitüntető Spinozát – elsősorban ez érdekli, hiszen „az emberi test, ha megmarad is a vérkeringés, s néhány más dolog, aminél fogva élőnek hisszük, mégis a magától egészen különböző természetre változhat” (*Etika* IV. 39. megj.). Az *Etika* gyógyászati célja e tekintetben a vérkeringésnél kifinomultabb és rafináltabb testi mozgásminták újraszabályozása lesz.

5. Tekintsük az államot, az *államtestet* (*imperii corpus*), ahogyan Spinoza fogalmaz a *Politikai tanulmány*ban. Ennek a testnek a részei az alattvalók (elsősorban emberek, másodsorban bármilyen egyéb individuum: állatok, ingóságok, ingatlanok és mindaz, amire az uralom kiterjedhet), sőt, még helyesebb úgy fogalmaznunk, hogy olyan alattvalók, akik már maguk is nagyobb szervekbe, testületekbe, céhekbe tömörülnek, melyek ugyancsak individuumok. Az államtest részei – tehát az alattvalók és a testületek – mozgásokat közölnek egymással, melyek összességükben más-más dinamikai szabálynak engedelmessé válnak a különböző berendezkedések alatt. A *Politikai tanulmány* – Spinoza érdeklődésének megfelelően – a berendezkedésekre mint alkotmányokra (demokratikus, arisztokratikus és monarchikus alkotmányokra) tekint, ám elvileg semmi akadálya annak, hogy tágítsuk a kört, és azt mondjuk, az államtest dinamikai szabályát az alkotmányos és törvényi szabályozásokon túl a társas érintkezés írott és íratlan szabályainak, a munkaidő és pihenőidő, a járműforgalom, áruforgalom, behozatal és kivitel, tőkeáramlás és minden egyéb mozgással, mozgatóssal kapcsolatos szabály összessége adja, amit éppoly képtelenek volnánk egyetlen formulába sűríteni, ahogyan ez már az emberi test dinamikai szabályánál is reménytelen vállalkozásnak tűnt. Ettől függetlenül a szabály érvényesül, ha pedig valamiért hatályát veszti, az államtest szétesik különálló individuumokra, és a rendszer átmegy természeti állapotba, vagyis a kölcsönös háborúskodás állapotába. Nincs kétség afelől, hogy Spinoza szemében a mozgásközlési rendszer ugyanolyan hierarchikusan szervezett az államtest esetében, mint az emberi test esetében, amit jól mutatnak azok a sorok, ahol Spinoza a holland állam bukása kapcsán az alkotmányról a *corporis humani fabrica* terminusaiban beszél: „a hollandok elegendőnek tartották a grófot letenni és az állam testét fejtől megfosztani, de nem gondoltak az alkotmány módosítására, hanem az államtest minden tagját meghagyták úgy, ahogy előbb volt, ekként megmaradt a hollandiai grófság gróf nélkül, mint egy test fej nélkül, s maga az állam elnevezés nélkül” (*Politikai tanulmány* IX. fej. 14. §). A testekről szerzett eddigi ismereteink alapján hozzátelhetjük ehhez a diagnózishoz, hogy nem a fej elvesztése miatt, hanem a test konstitúciójának, következképpen a dinamika szabályának a változatlanul hagyása miatt bukott el a holland állam: az államtest fej nélkül is működőképes lehet, ám tudomásul kell venni, hogy ekkor már egy új test, új hierarchiával, amely egy új dinamikai szabállyal marad csak működőképes. Az „elnevezés nélküli”

állam egyfajta kiméra: olyan, mintha egy oroszlán egy sas dinamikai szabálya szerint próbálna meg létezni – vadászni, repülni, fészket rakni.

6. Végül pedig ott van a természet egésze, mint egyetlen természettest: „az egész természet egyetlen individuum, melynek részei, azaz valamennyi test, végtelen sok módon változnak anélkül, hogy az egész individuum bármiképp megváltoznék” (*Etika* II. 7. segédétel megj.). Ez a test semmilyen általunk ismert testre nem hasonlít. A természet nem egy szilárd, folyékony vagy légneemű test, nem is organizmus, és nem is egy állam (amit Isten kormányzóként igazgatna). A természet dinamikai szabálya ezen individuumokétól eltérő, náluk mérhetetlenül összetettebb szabály, amiről Spinoza csak egyetlen konkrétumot tud állítani: azt, hogy ez a szabály nem változhat meg. Ha a természetben belüli individuumok szüntelenül átalakulnak más individuumokká, a természet maga mint individuum nem alakul át más individuummá, hanem ugyanaz marad. A természettől idegen a betegség és a halál, a forradalom és a háború: a természet rendje minden képzeletet felülmúló békességgel és háborítatlan nyugalommal érvényesül. Mivel pontosan ez a béke és nyugalom az, ami kifejeződik a bölcs ember lelki nyugalmában, ezért van a természet megismerésének oly kitüntetett etikai jelentősége Spinozánál.

### III. A TEST MINT PROBLÉMA

Miután tisztáztuk a spinozai megfogalmazások jelentését, a továbbiakban az egyszerűség kedvéért mind a bizonyos meghatározott aránynak engedelmessé váló mozgásközlésre, mind a mozgás–nyugalom, gyorsaság–lassúság arányára vagy viszonyára az illető test *dinamikai szabályaként* fogunk hivatkozni. Általánosságban kijelenthetjük, hogy a spinozizmusban minden testnek van egy dinamikai szabálya, amely a testet, az épp így létező testet konstituálja. Illetve megfordítva, ha a természetben valahol, valahány test, valamilyen dinamikai szabály szerint rendeződik el, úgy szükségképpen létrejön egy új test, amelynek pontosan ez a szabály lesz a dinamikai szabálya. Röviden tehát a test a spinozizmusban egy szabályos dinamikai rendszer. A szabálytalan dinamikai rendszerek nem testek. Spinoza kijelentését: „Persze, hogy mire képes a test, még senki sem határozta meg eddig...” (*Etika* III. 2. megj.), most már lefordíthatjuk úgy is, hogy a tudomány még nem tárta fel a természetben található különféle testek, és legfőképpen az élő szervezetek pontos dinamikai szabályait. Céloztunk már rá, hogy nemcsak a fizika, a matematika vagy az anatómia, de maga a spinozai filozófia sem állapít meg ilyen szabályokat (eltekintve persze a *Rövid tanulmány* 1:3 arányától és a *Politikai tanulmány* alkotmányformáitól mint konkrét „áttörésektől”), hanem leginkább csak tételezi őket. Tételezi őket mint *filozófiai fogalmat*, ami persze már önmagában is eredmény, hiszen a fogalom az, ami lehetővé teszi, hogy olyan dolgokat lássunk meg, amik nélküle láthatatlanok és elgondolha-

atlanok maradnának. Joggal jelenthetjük ki ezért, hogy a testnek mint szabályos dinamikai rendszernek a fogalma, mellyel a természet egészét egy egységes logika mentén tekinthetjük át, semmivel sem kevésbé egzakt fogalom, mint például a matematikai függvény fogalma, amely pontok és számok végtelen sokaságát teszi a hozzárendelési szabály egységes logikája mentén szemléletessé és áttekinthetővé.

A kérdés, ami ezek után önkéntelenül is fölvetődik, a következő: mi egy szabály? Mi különbözteti meg egymástól a szabályos és a szabálytalan dinamikai rendszereket, más szóval a testeket és a nem testeket? Mi is tehát alapjában véve a test?

Ezt a kérdést pedzegettük már a *causa seu ratio* és az inkorporáló kauzalitás két perspektívájának megkülönböztetésekor is. Ehhez hozzátettük akkor, hogy Spinozánál az első uralja a másodikat, amelyik implicit marad, továbbá mindkettő komplementer problematizációs utakat nyit meg. Nézzük most meg közelebbről, hogy milyenek ezek a problematizációs utak. A *causa seu ratio* explicit perspektívájánál legegyszerűbb lesz, ha Gilles Deleuze-höz fordulunk, aki tömören és világosan összefoglalta az e perspektívából nyíló problematizációs csapást.<sup>11</sup> A probléma a következő: adva van egy test – egy pók, egy ló, egy gyermek, egy iszákos, egy skizofrén teste, vagy akár az a test, ami a sajátom –, és keresendő az a karakterisztikus arány, az a dinamikai szabály, amely a természetét meghatározza, a test ilyenségének az oka és alapja. Deleuze szerint ez a spinozai probléma egyszerre fizikai, etikai, etológiai, gyógyászati és politikai probléma. Hiszen a dinamikai szabály meghatározza, hogy a test mit képes elviselni és mit nem, milyen affektusokra hajlamos és milyenekre nem, milyen hatásokat produkálhat és milyeneket nem, mi ártalmas számára és mi növeli az egészségét, milyen testekkel alkothat közösséget és milyenekkel nem. „Mond meg, milyen affekciókra vagy képes, s megmondom, ki vagy!” – hozza spinozista formára Deleuze a *gnóthi szeauton* ókori eszméjét. A dinamikai szabálynak és a belőle következő faktoroknak a feltérképezése – kartográfiája – egyet tesz a spinozai *Etika* gyakorlatba való átültetésével. A *causa seu ratio* problematizációs csapása egyszóval a spinozai etika életvitelszerű gyakorlata.

Az inkorporáló kauzalitás perspektívájából nyíló problematizációs csapás az előző mellé állítva a következőképpen fest: adva van egy mozgásközlési rendszer, sok-sok egyszerre cselekvő test, és keresendő újfent a karakterisztikus arány, az a dinamikai szabály, amelynek jóvoltából e sok test egyetlen komplex testté individuálódik. A probléma tehát mindkét esetben ugyanaz: maga a szabály, az individuáló szabály. Épp csak a második esetben semmi biztosítékunk nincs arra, hogy létezik szabály, azaz létezik szabálynak megfelelő test.

De miért fontos kérdés ez? – vethetnénk itt közbe joggal. Azt láttuk, hogy az első problematizációs csapásnak milyen praktikus jelentősége van. Ám mi a je-

<sup>11</sup> Lásd Deleuze 1980; 1974. 01. 14 és 1977. 02. 15.

lentsége ennek a másodiknak? Végző soron nem mindegy, hogy testnek vagy nem testnek tartjuk azt, ami egy mozgásközlési rendszerben kibontakozik? Látni fogjuk, hogy nem, méghozzá több okból sem.

Kezdjük azzal, hogy az inkorporáló kauzalitás perspektívájából feltett kérdéseink rávilágítanak egy spinozai paradoxonra, amelyet így foglalhatnánk össze: a természet egésze egyetlen individuum, egyetlen test, amely minden létező testet magában foglal, ennek dacára mégsem állíthatjuk azt, hogy ennek az egyetlen testnek bármely tetszőleges részhalmaza *testet* alkotna. A természet minden teste egyetlen cselekvésben összehatva létrehozza a természettestet, de ez a fa a kertemben, az az ember Ázsiában, az a kő a Holdon – miközben részt vesznek ebben az egyetlen cselekvésben – mégsem alkotnak együttvéve egy testet. Nem, hiszen nem váltanak ki közösen okozatot, és nincsen közös dinamikai szabályuk sem, amelyre hivatkozva egyetlen individuumnak, egyetlen testnek tekinthetnénk őket.

Ezt a paradoxont átfogalmazhatjuk a következő kérdés formájában: vajon megkülönböztethető-e Spinozánál egy egyedi test a körülmények összejátszásától? Vizsgáljuk meg azt a példát, amit Spinoza az *Etika* I. részének *Függelékében* hoz fel: egy ember, akit barátja vendégségbe hívott, a barátja háza felé tart az utcán, amikor a szél fölkap egy tetőcserepet, ami az illető fejére hullik, aki szörnyethal. A babonások isteni cselekvést látnának ebben, hiszen máskülönben – tesz fel a kérdést – „hogyan hathatott volna véletlenül össze ennyi körülmény (mert gyakran sok körülmény hat egyszerre)”. Gyakran sok körülmény hat egyszerre, írja tehát Spinoza, aki egyébként ebben az eseményben nem isteni beavatkozást lát, hanem a körülmények összejátszását: *multae simul concurrunt*. Ha visszaemlékszünk az egyedi dolog definíciójára, ott is valami hasonlót olvastunk: *in una actione concurrant*, sok individuum hat össze egy cselekvésben. A kérdés tehát, ami most már a spinozai szövegeken alapul, a következő: vajon van-e *elvi* különbség a két összehatás között? Van-e *elvi* akadály annak, hogy a körülményeket egyetlen testnek tekintsük, a testeket, az egyedi dolgokat pedig körülményeknek? Persze e konkrét példánál hajlunk arra, hogy ne tekintsük az eseményt – benne a barátot, az emberünket, a szelet és a tetőcserepet – egyetlen testnek, hanem inkább csak két mozgásközlési rendszer összetalálkozásának. Egyrészt a barát meghívja emberünket, s ezzel mozgásba hozza őt; másrészt a szél, vagyis a levegőt alkotó individuumok olyképpen hatnak össze, hogy egyetlen okozatot kiváltva elmozdítják a tetőcserepet, amely összetalálkozik a mozgásba hozott emberünkkel, akinek átadja a maga igen heves mozgását. Ám nem látjuk, hogy milyen elvi akadály lenne annak, hogy e testeket egyetlen testnek tekintsük – hiszen ahogyan azt állítottuk, hogy nem igaz az, hogy a természet bármely tetszőleges részhalmaza testet alkot, úgy másfelől az is igaz, hogy a természet bármely tetszőleges részhalmaza alkalomadtán egy testté *szerveződhet*. Innentől kezdve paradoxonunk átalakul egy gyakorlati, már-már művészeti követelménnyé: észre kell tudnunk venni a különbséget a körülmények és a tes-

tek között, és észre kell tudnunk venni, ha a körülményekben inkorporálódik egy test. Hiszen ha *elvéleg* nincsen különbség körülmény és test közt, úgy mindig *művészet* lesz meglátni a különbséget köztük.

Hogy ennek a művészetnek milyen gyakorlati jelentősége van, azt két példán mutatjuk meg. Egyrészt már korábban megfogalmaztuk a fenti paradoxonnak egy másik aspektusát, mint a kimératest és a valódi test megkülönböztetésének a problémáját. Vajon meg tudunk-e különböztetni egy kimératestet egy valódi testtől? – kérdeztük akkor. Valójában időközben már talákoztunk egy olyan példával, amikor pontosan ez volt a probléma, még hozzá egy politikai természetű, gyakorlati probléma: a holland állam bukása a *Politikai tanulmányban*. A holland állam azért bukott meg, mert akik a politikát irányították, megfosztották az állam testét a fejétől, ám nem gondoltak az alkotmány – vagyis, tettük hozzá, az államtest dinamikai szabályának – módosítására. Magyarán a holland politika irányítói nem vették észre, hogy a megváltozott körülmények között az alattvalók és az intézmények inkorporáló kauzalitása egy kimératestet, és nem valódi testet eredményez. Nem vették észre a kimérát. A szabály félreismerése pedig egy állam bukásához vezetett. Általánosságban kijelenthetjük, hogy az inkorporáló kauzalitás problémája a politika terén a politikai szereplők szempontjából vetődik fel, és így hangzik: az individuumok aktuális összjátéka vajon milyen szabályokat követ? Állam-e még ez az összjáték? Demokrácia-e még? Testet alkot-e még? És nem inkorporálódik-e az állam leple alatt egy forradalom vagy lázadás teste, amely az államot szétvetheti, és aminek a szabályát ha nem bogyozzuk ki, az államtest szétesését csak a körülmények összjátékának tudhatjuk majd be?

A holland állam bukásának van egy másik figyelemre méltó tanulsága is. Az, hogy a *causa seu ratio* perspektívájával szemben az inkorporáló kauzalitás perspektívájában igenis kijelenthetjük: a természet eltűri a kimérákat. Bár csak ideig-óráig, de eltűri őket. A holland államkiméra ideig-óráig fennáll, majd elbukik, mert nem életképes. Igazat adhatunk Zourabichvilinek, aki a spinozai kimérát úgy határozza meg, mint „strukturálisan önpusztító” létezőt, amely felszámolja magát, mert a természete ellentmondást hordoz magában, mert a testét konstituáló dinamikai szabály olyan, hogy önnön részeit egymás ellen fordítja (Zourabichvili 2002. 221). Ám a természet eltűri az önpusztító embert, eltűri az öngyilkost, addig legalábbis, amíg teljesen fel nem számolja magát. Innentől kezdve a kiméra kérdése már nem csupán egy ismeretelméleti kérdés (milyen módszerrel tudjuk felismerni és gondolkodásunkból eltávolítani a kimérákat?), hanem az a hús-vér fizikai kérdés, hogy meddig tart egy rendszer nem kimérikus állapota, hol húzódik a határ, amit átlépve kimérává változik, s így megjósolhatóan gyors véget ér.

A szabály kibogozásának művészetére felhozható másik példánkat röviden a szimbiózis problémájának nevezhetjük, ami Spinozánál a közösségalkotás címszava alatt kerül elő. A természetben bármely két test közösségbe kovácsolódhat – ha csak ideig-óráig is – egy meghatározott dinamikai szabály szerint. Ez a

szabály lehet vagy a két test egyikének a szabálya, vagy egy olyan közös szabály, amelynek a két test szabálya alárendelődik.

Az első esetnek két változata lehet: (1) Egy test rákényszerítheti úgy a szabályát egy másik testre, hogy az elpusztul. Például táplálkozásnál a táplálék bekebelezésével és megemésztésével a táplálkozó test rákényszeríti a saját testének dinamikai szabályát a tápláléktestre, melynek alkotórészei kiszabadulnak eddigi dinamikai szabályuk alól, és a táplálkozó test dinamikai szabálya szerint rendeződnek el. Ha ellenben a táplálék méregnek bizonyul, akkor a méreg fogja a saját dinamikai szabálya alá vonni a táplálkozó test részeit, és elpusztítja azt. (2) Egy test rákényszerítheti a szabályát egy másik testre úgy is, hogy az nem pusztul el, hanem annak részévé válik. Például az ember része lehet egy államnak, majd letelepedhet egy másik államban is, és mindkét esetben ennek vagy annak az államtestnek az ilyen vagy olyan dinamikai szabálya fogja felüldeterminálni a dinamikai szabályát anélkül, hogy elpusztítaná az illetőt.

A második esetben, amikor tehát két test dinamikai szabálya egyetlen közös dinamikai szabály alá rendelődik, alkalmazhatjuk az individuум definícióját, mely szerint ekkor szükségképpen létrejön egy új test, amelynek az első kettő része lesz. Ez az új test addig áll fenn, amíg a közös szabály hatályát nem veszti, s ekkor az individuум felbomlik azokra az összetevő individuúmokra, amelyekből eredetileg létrejött. Spinoza klasszikus példája erre a nyirok és a hilus, mint két önálló dinamikai szabállyal rendelkező individuум, melyek egy közös szabály alatt egyetlen individuummá, tudniillik vérré egyesülnek.<sup>12</sup> Ez a második eset az, amit röviden szimbiózisnak nevezhetünk. Ilyen szimbiózist alkot egy futbalcsapat, egy zenekar, egy állam polgárai, egy ökológiai milió stb.

Előre sejtethetjük ezek után, hogy miként hangzik majd a szimbiózis problémája az inkorporáló kauzalitás perspektívájában. Adva van két (vagy több) test, melyek szimbiózisban élnek egymással, és keresendő az a szabály, amely a szimbiózist megszervezi, illetve keresendő az az új, nagyobb test, amely a szimbiózisban létrejön. A dohányos ember és a nikotin, az alkoholista és az alkohol, a kábítószeres és a drog, az élő szervezet és vírusai – a sort hosszasan folytathatnánk – mind a szimbiózisnak ezt a problémáját vetik fel. Az is sejtethető az előzőekből, hogy itt is szembe kell nézni azzal az eshetőséggel, hogy a szimbiózisban létrejövő test csupán egy életképtelen kimératest. Ennek az eshetőségnek ráadásul annál is nagyobb a jelentősége, mivel a kimératest dinamikai szabálya, mint láttuk, egymás ellen fordíthatja önnön részeit. Így a kimératest megszűnésével nemcsak maga ez a test szűnhet meg, hanem az alkotórészei is széteshetnek. Hiszen míg az életképes szimbiózis, például a vér esetében, a vér megsemmisülésével a hilus és a nyirok sértetlenül megmarad, addig egy kiméraszerű szimbiózis esetében, például egy alkoholos vagy kábítószeres szervezet esetében a szimbiózistest megsemmisülését nem feltétlenül éli túl a szimbió-

<sup>12</sup> Lásd Spinoza 32. levelét Oldenburghoz.

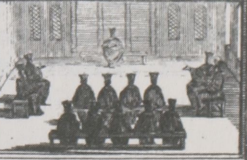
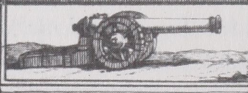
ziséből kiszakadó szervezet. A kiméraszerű szimbiózis, így vagy úgy, pusztulásra ítéli a tagjait.

Az *Étika* az ész által vezetett emberek szimbiózisával egy olyan hatalmasabb test kialakulását ígéri, amely az egyes tagjainak az erejét és hatalmát is növeli. A test problémájának eddigi elemzései azt mutatják – és ezzel a „szocio-spinozizmus” számára is levonhatjuk a tanulságot –, hogy ez nem minden szimbiózis esetében van így. Vannak olyan szimbiózisok, melyek úgy válnak hatalmasabb testté, hogy közben élőködnek tagjaikon, és kimérákká változtatják őket. Úgy tűnik, hamarosan igen aktuálissá válhat ez a kérdés egy másik területen is, nevezetesen az emberi test és a technológiai test biotechnológiai szimbiózisánál: Vajon fel tudjuk-e ismerni idejében a kimérát?

#### IRODALOM

- Citton, Yves – Frédéric Lordon (szerk.) 2008. *Spinoza et les sciences sociales. De l'économie des affects à la puissance de la multitude*. Paris, Éditions Amsterdam.
- Deleuze, Gilles 1980. *Mille plateaux*. Paris, Minuit.
- Deleuze, Gilles *Cours de Vincennes*, 1974. 01. 14.; 1977. 02. 15.; 1981. 02. 17. [www.webdeleuze.com](http://www.webdeleuze.com) [A hozzáférés időpontja: 2010. május 31.]
- Descartes, René 1647/1997. *Les Principes de la Philosophie*. In *Ű Oeuvres philosophiques*. Szerk. F. Alquié. III. kötet. Paris, Garnier.
- Spinoza, Benedictus 1677/1997. *Étika*. Szemere Samu fordítását átdolgozta, a bevezetőt, az utószót, a jegyzeteket írta és a mutatót összeállította Boros Gábor. Budapest, Osiris.
- Spinoza, Benedictus 1980. *Politikai tanulmány és levelezés*. Ford. Szemere Samu. Budapest, Akadémiai Kiadó.
- Spinoza, Benedictus 1661/1981. *Rövid tanulmány*. In *Ű Ifjúkori művek*. Ford. Szemere Samu. Budapest, Akadémiai Kiadó.
- Zourabichvili, François 2002. *Spinoza. Une physique de la pensée*. Paris, PUF.

*Non est potestas Super Terram quae Comparatur ei Job 41 27*



# LEVIATHAN

Or  
THE MATTER, FORME  
and POWER of A COMMON-  
WEALTH ECCLESIASTICALL  
and CIVIL.

By THOMAS HOBBS  
of MALMESBVRY.

London  
Printed for Andrew Crooke  
1652



## A Leviatán anatómiája

Az uralom minden formája szorosan kötődik azokhoz a képi és retorikai reprezentációkhoz, illetve médiumaikhoz, amelyekben keresztül a nyilvánosság előtt megjelenik. Ez a reprezentáció, amely elsősorban az uralom intézményeihez, az erőszak gyakorlásának módjaihoz, illetve a politikai közösséghez kötődik, mindig tartalmaz a politika imaginárius szférájára való utalást, ahol az uralom a végső alakját és igazolását elnyeri. A politika ebben az értelemben nem egyszerűen a fizikai erőszak legitím és közvetlen gyakorlása, vagy funkcionális differenciálódás révén kialakult intézményrendszer, de nem is a politikai együttélés kerete, hanem egy olyan rend, amely közvetítve vagy metaforikusan jelenik meg. Ez a képi és retorikai reprezentáció azonban nem egyszerűen az uralmat, illetve annak eszközeit és intézményeit közvetíti, hanem egy alapvetően imaginárius politikai szférát konstruál, amely maga is meghatározza az uralom jellegét (vö. Maye 2007. 15). Ez teremti meg a politikai együttélés alapjait és a hozzá kapcsolódó uralom sajátos rendjét. Minden politikai közösség egy fiktív aktusból vagy mítoszból születik, amelyek az alapítások vagy az újraalapítások emlékeit őrzik. Ezeket állandóan újra mesélik, vagy képileg megjelenítik, mert a politikai mítosz az, amelyen keresztül a mindenkori jelen a hagyományhoz kapcsolódik, ami azonban nem azonos a múlttal. Az eredetmítoszok az alapításokhoz, a modern mítoszok a fiktív szerződéskötésekhez és forradalmi cezúrákhoz kötődnek. Az alapítások és újraalapítások eredendően olyan események, amelyek politikai mítoszokként kontinuitásokat és diszkontinuitásokat teremtenek, kiválogatva és osztályozva a múltból kiemelendő és feledésre ítélt eseményeket, a történeti és fiktív elemeket elegyítve. Ennek révén olyan identitást teremtenek, amely egyfajta önképet alkotva egyrészt eltünteti a társadalmi és kulturális különbségeket, másrészt állandóan újrafogalmazza a társadalmi rendet és hierarchiát. A folyamat során a politika imaginárius szférája folyamatosan reaktualizálódik. Minden politikai jelenség referenciális alapja erre a szférára vezethető vissza, és a politika valamennyi szimbolikus eleme ehhez viszonyul. Minden szimbolikus politikai eselekvés, rítus, jelvény, szimbólum erre az imaginárius politikai szférára utal vissza, itt vannak a gyökerei, ebből újul meg, vagy éppen ezzel szakít.

Ezek megértéséhez, értelmezéséhez egy sor genuin esztétikai folyamatra van szükség, amelyek egyszerre játszanak szerepet a politika közvetlen inszenzációjában és recepciójában (vö. Balke–Scholz 2007. 35).

Ezen esztétikai procedúrák közvetítésének talán legideálisabb, de minden bizonnyal legerjedtebb médiuma az emberi test, amelyben a legkülönbözőbb metaforikus jelentések sűrűsödnek össze, s ezért a legváltozatosabb formákban ölthetnek testet. A test különösen alkalmas az imaginárius politikai terek megjelenítésére és az uralom fizikai dimenzióinak közvetítésére. Elsősorban annak köszönheti e kitüntetett szerepét, hogy egyszerre természeti és kulturális konstrukció; egyszerre organikus szervezet és mechanikus szerkezet, és mivel e vonásai részben ellentétes, részben komplementer viszonyban állnak egymással, így állandó kölcsönhatásban vannak. A test egyszerre partikuláris és univerzális, szent és profán, egységes és tagolt, élő és növekvő organizmus, illetve éles határokkal rendelkező tárgy, amely így minden kontinuitás és diszkontinuitás hordozója egyben. A test egyszerre tartalmazza a változékonyt és az állandót. Egyszerre jelent természeti adottságot vagy isteni adományt, illetve emberi alkotást; egyrészt adományként kapjuk, másrészt ápoljuk, építjük, gyógyítjuk. A test isteni és emberi; egyszerre jelenti az ember teremtőképességének kiszélesedését, amellyel az isteni teremtés helyébe lép, és sebezhetőségét, amely mindig véges mivoltára emlékezteti. A test az emberi képességek szinte korlátlan fokozásának és örök korlátozottságának helye. Egyszerre van benne az ember halandósága és halhatatlansága, végessége és végtelensége. A testben egyszerre van jelen az uralom és az uralhatatlanság. Éppen ezért a testtel kapcsolatban nem alkalmazható a *phüszisz* és *nomosz* közötti megkülönböztetés, mivel egyik sem lehet kizárólag, és a kettő között kölcsönös függőség áll fenn. A természet maga sem változatlan, és a természet változásait maga az ember próbálja meg befolyásolni, ezért lehetetlen élesen megkülönböztetni a testtel kapcsolatos folyamatokat. A hatalmi viszonyok és rendelkezései különösen plasztikusan jelennek meg a testben, aminek révén az primér politikai jelentőségre tesz szert. A testben mint az állam vagy a közösség metaforájában mindig a világ rendje testesült meg, úgy tekintettek rá, mint magára a kozmoszra kicsiben. Mivel az emberi közösségek nem rendelkeznek valódi fizikai határokkal, ezért a test analógiája magától értetődő volt mind a politikai közösség egysége, mind az uralom gyakorlásának helye szempontjából. A politikai közösséget gyakran úgy írták le, mint természetes vagy mesterséges testet, amely bizonyos társadalmi helyzeteknek és funkcióknak (mint Titus Liviusnál Menenius Agrippa meséjében) vagy éppen erkölcsi értékeknek (mint John of Salisbury *Policriticus*ában) megfelelő (test)részekből áll, mégis képes az egység kifejezésére. A test elsősorban azért ideális az imaginárius politikai szféra és a politika uralmi tere közötti metaforikus közvetítés szerepére, mert képes a rendet az imaginárius szférából az uralom fizikai terébe közvetíteni.

## I. A LEVIATÁN BONCOLÁSA

Valamennyi politikai test közül a legnagyobb hatású Hobbes Leviatánja, amely úgy tűnik, képes egyesíteni az imaginárius politikai térrel kapcsolatos elképzeléseket. A Leviatán metaforája nem a valóság reprezentációja, hanem olyan összefüggések és értelmezések egész sora, amelyek a metaforát képpé, a képet metaforává alakítják. Ennek következtében a Leviatán nem egyszerűen a szuverén és rajta keresztül az állam leghatásosabb szimbóluma, hanem egy sokrétegű jelentéssz összefüggés hordozója. Éppen ebben a sokoldalúságban rejlik annak az alapvetően ellentmondásos jellegnek a magyarázata, amely a Leviatán képi és retorikai metaforáját keletkezésétől fogva körülveszi. A kép ugyanis ebben az esetben nem egyszerűen másodlagos illusztráció a szöveghez képest. Minél több médiuma van egy metaforának, annál komplexebb értelmezési minták rajzolódnak ki, és így már nem a kép valóságtartalma a fontos, hanem a hatása, amelynek következtében egyre inkább távolodunk szerzőjének eredeti megfontolásaitól.

Thomas Hobbes *Leviatán, vagy az egyházi és világi állam formája és hatalma* című műve 1651-es kiadásának címlapját Abraham Bosse készítette Párizsban, valószínűleg a szerző közreműködésével.<sup>1</sup> A kép horizontálisan két nagy részre osztható. A felső képen az óriás a szinte láthatatlan tengerből emelkedik ki, a dombok fölé, amelynek előterében egy város látható. Az óriás koronát visel, jobb kezével kardot emel, a bal kezében pedig püspöki pásztorbotot tart, amelyeknek árnyéka a földre vetődik. Ezek utalnak a könyv alcímében említett „egyházi és világi állam formájára és hatalmára”. A kard és a pásztorbot egy háromszöget zárnak be, amelyet felül a Jób könyvéből származó idézet zár le: *Non est potestas Super Terram quae Comparetur ei* – „Nincs e földön hozzá hasonló hatalom” (Jób 41, 33).<sup>2</sup> Az óriás teste mintegy háromszáz emberi alakból áll, akik valamennyien a nézőnek háttal állnak, tekintetüket a fej irányába fordítják. Ezek az alakok alkotják az óriás testét, a fej és a kezek kivételével, amelyek így némileg elkülönülnek. A kép alsó fele három részből áll. Az oldalsó képsorok a kard és a püspöki pásztorbot szembeállításának meghosszabbításai. A bal oldali mezők a világi, a jobb oldaliak az egyházi hatalmat jelképezik, és a könyv második, illetve harmadik részének felelnek meg. A két oszlop mezői horizontálisan is össze vannak kötve: a vár a templommal, a korona a süveggel, az ágyú a kiátkozás villámaival, a harci jelvények a logika fegyvereivel, a csata a hitvitákkal. A mezők egy közbeeső területet szegélyeznek, amelyet egy függöny takar; ennek felirata

<sup>1</sup> A *Leviatán* címlapját eredetileg Wenceslaus (Václav) Hollarnak tulajdonították, újabban azonban Horst Bredekamp kutatásai alapján egyre biztosabbnak tűnik, hogy az alkotó Abraham Bosse (1604–1676), aki korának híres rézmetszője és a párizsi Királyi Akadémia (Académie Royale de Peinture et de Sculpture) tanára, majd tagja volt. Vö. Bredekamp 2003, 39–52.

<sup>2</sup> A bibliai idézetek a tanulmányban a Hobbes által használt latin verzió fordításai.

a könyv címét, szerzőjét, illetve a kiadás helyére és idejére vonatkozó adatokat tartalmazza.

Ha közelebbről szemügyre vesszük a *Leviatán* alakját, akkor egy olyan sokrétegű kép tárul elénk, amelyben keverednek a racionális számítások és a félelem kiváltotta fantáziák. Ennek nyomán hamar kiderül, hogy nem egyetlen *Leviatán* van dolgunk, hanem egy köztes lényel, amely – ellentétben az ókor félisteneivel, akiknek lakhelye az ismert világ vége vagy a túlvilág volt – a valóság és a fikció között helyezkedik el. Kétségtelen, hogy a *Leviatán* nem sorolható be rendszerint, mivel az isteni, az emberi, az állati és a mechanikus közötti szűrkezőnában található, és e vonások csak bizonyos értelmi összefüggésekben és politikai kondíciók között nyerik el saját jelentésüket. De vajon mi sűrűsíthető bele a *Leviatán* metaforájába, és milyen alakzatok kristályosodnak ki alakjában? Ez a kérdés azért különösen érdekes, mert Hobbes ugyan igyekszik kerülni az organikus metaforák nyelvét, és elsősorban a racionális érvelésre támaszkodni (vö. Ludassy 207. 46), a *Leviatán* azonban már önmagában is metaforákat hordoz, amelyek folyamatosan újabb és újabb tartalmakkal bővülnek. Úgy tűnik, hogy a logikai érvek megbicsaklása és a metaforikus bővülés közötti összefüggés csak részben ad választ arra a zavarra, ami a *Leviatán* alakját kezdettől fogva körülveszi.

A *Leviatán* egy olyan köztes lény, amelyről csak akkor tudjuk pontosan megállapítani, hogy mit jelent, illetve mi a státusza, ha a szó szoros értelmében anatómiai vizsgálatnak vetjük alá, azaz felboncoljuk alakjai és természetei szerint. Ennek alapján a *Leviatán*ban legalább négy metafora: a mechanikus szerkezet, a halandó isten, a mesterséges ember és a szörnyszerű állat, illetve ezek alapvető karakterei – az instrumentális, az exkluzív, az antropomorf és a mitikus – sűrűsödnek össze részben kiegészítve, részben árnyalva egymás jelentését.

### *1. A mechanikus gépezet*

Hobbes a következőket írja a *Leviatán* Bevezetésében:

Ha ugyanis felismerjük, hogy az élet nem egyéb, mint végtagok mozgása, amelynek kiindulási helye valamilyen fontos belső szerv, akkor miért ne állíthatnánk, hogy minden automatának (olyan szerkezetnek, amely – mint például az óra – rugók és kerekek segítségével önmagától működik) mesterséges élete van? Hisz mi más a szív, mint rugó, s mi mások az idegek, mint megannyi huzal, s az ízületek, mint megannyi kerék, amelyek az egész testet az Alkotó kívánsága szerint mozgatják? (Hobbes 1999. I. 69)

Az újkor kezdetén az emberi közösségek hosszú hagyományra visszatekintő organikus képzete fokozatosan eltűnik, és helyét egyre inkább a társadalom mechanikus felfogása foglalja el, ami elsősorban az ember teremtőképességének

kiszélesedésével és a különböző gazdasági és politikai funkciók differenciálódásával áll összefüggésben. A politikai és társadalmi gépmetaforák ebben a korban elsősorban az óraművekre utalnak, amelyeken keresztül a világ mint egy mechanikus szerkezet, Isten pedig mint órásmeister tűnik fel. A kozmosz analogonja az ember által készített szerkezet, mintegy a teremtés megismétlése, amelynek során az ember Isten helyére lép. Az ember számára ennek egyetlen feltétele az isteni eredetű természeti törvények mechanikai elvekre való lefordítása és technikai alkalmazása. Ezekben a törvényeken túl az emberi képzeletnek semmi sem szab határt. A racionalitás ebben a tekintetben az isteni törvény másik arca, amelynek révén a valóság matematikailag leírható és így újrateremhető. Mivel minden e törvényszerűségeknek engedelmeskedik, ez feltételezni engedi, hogy az ezt veszélyeztető emberi kontingenciát ki lehet iktatni, és az embert úgy lehet alakítani, hogy fogaskerékként tökéletesen együttműködjön másokkal, és lehetővé tegye a pontos és tervszerű működést, amelyet immár semmi sem zavarhat. A rész aláveti magát az egésznek, mert így ő maga is egy magasabb rendű cél elemévé válik, ami a maga politikai vonatkozásait tekintve az államgépezet fogalmában ölt testet. Hobbes a Leviatán gépezetként való felfogásával elsősorban annak alapvető instrumentális karakterére utal, és egy cél elérését szolgáló eszköznek tekinti, ami csak addig marad fenn, amíg célját, az emberek közötti békét és biztonságot képes szavatolni, jóllehet végcélját sohasem képes maradéktalanul elérni. A cél a szó legszorosabb értelmében szentesíti az eszközt, és ennek fényében módosul, illetve relativizálódik, amennyiben már nem a békével, hanem a túléléssel azonos.

## 2. A halandó Isten

Ha a Leviatán a halandó Isten, aki a halhatatlan Isten fennhatósága alatt áll, akkor úgy tekinthetjük, hogy a halandó Isten a halhatatlan Isten helyére lép a Földön. Ha mindez igaz, akkor a halandó Isten országa a természeti állapot által megjelenített gonosszal áll szemben. A halandó Istent emberek alkotják, ugyanakkor, mivel a fej és a kézfejek szabadon maradnak, és kizárólag ezek viselnek jelvényeket – koronát, kardot, püspöki pásztorbotot –, el is különülnek. Az ember a teremtés aktusával egyben maga fölé is helyezi a halandó Istent, aki így függetlenné válik, uralmának megalapozásához nincs szüksége a halhatatlan Isten segítségére. *A Jób könyvéből* származó idézet, amely az óriás feje fölött lezárja a teret, kizárja a magasabb rendű legitimáció lehetőségét és szükségességét, miközben arra figyelmeztet, hogy nincs hozzá fogható hatalmú és erejű jelenség. A halandó Isten csak a halhatatlan Istennek tartozik felelősséggel, de uralma egyáltalán nem függ tőle. Hobbes meg is tesz mindent, hogy a teodícea problémáját kiküszöbölje azzal, hogy a halandó és halhatatlan Istent határozottan elválasztja egymástól, és amíg az előbbit kizárólagossá és kétségbevonhatat-

lanná teszi, addig az utóbbit teljesen kívül helyezi a képletben, csupán a végső ítélő szerepét szánja neki. A Leviatán Istene politikailag halott, és a szuverén uralma csak úgy képzelhető el, ha Isten helye a képletben üres marad.

Ez a körülmény azt a vonást hangsúlyozza, hogy a Leviatán semmilyen ellentmondást nem tűr, és bármennyire ember alkotta lénynek tekinthető, lehetetlen vele szembeszállni. A Leviatán isteni jellegéből fakadó érinthetetlensége arra utal, hogy Hobbes egy alapvetően profán politikai jelenséget szakralizál, ezzel együtt azonban igyekszik teremtményét az idolátria minden gyanújától megszabadítani. A politikai jelenségek szakralizálása éppen azt a célt szolgálja, hogy az olyan intézményeket és eszközöket, mint az állam és a hozzá kapcsolódó erőszak, egyrészt igazolja, másrészt egy magasabb rendű, de alapjában véve mégis evilági cél végrehajtóinak tekintse. Ennek jogosságához így semmilyen kétség nem férhet, és a vele való bármilyen szembefordulás rögvest kiátkozást vagy megtorlást vált ki. „Az állam az egyház, és aki az új *ecclesia*, vagyis a Leviatán ellenségeként lép fel, az a Sátán” (Voegelin 1938/1996. 48). Mindazonáltal a *deus mortalis* metaforájában végképp összekeverednek az isteni és emberi, a halhatatlan és halandó vonások, amelyek a tartósan stabilizálni kívánt biztonság és béke állapota, illetve az Isten halandósága kapcsán nyilvánvalóan ellentmondásba kerülnek.

### 3. A mesterséges ember

A Leviatán mint *homo magnus* teste emberi alakokból épül fel, ami nyilván arra utal, hogy az emberek a Leviatán testében egyesülnek. A címlapon egyetlen ember sem látható a vidéken vagy a városban, mindannyian egyetlen testet alkotnak, ami a szerződéskötés egyesítő pillanatára utal.

Instituált államról akkor beszélünk, ha egy emberi sokaságban mindenki mindenkivel megállapodás útján megegyezik abban, hogy a többség mindannyiuk megszemélyesítésének, azaz képviselőjének jogát bárkire vagy bármilyen gyülekezetre ruházza át; [...] az emberek a megszemélyesítés jogát arra, akit uralkodójuknak neveznek, csak egymás közt, nem pedig mindannyiuk és közte létrejött szerződés alapján ruházták át. [...] Következésképpen egyetlen alattvalója sem szabadíthatja ki magát az alárendeltségből azáltal, hogy e szerződés érvénytelenségével hozakodik elő (Hobbes 1999. 210–211).

A szerződéskötés kizárólag az embereket kötelezi, egymással kötnek szerződést és nem a szuverénnel. Közülük bárki lehet a szuverén – jöllehet csak egyikük. Ez a körülmény minden ember születésétől fogva adott egyenlőségére utal, ami azonban csak a szerződéskötés aktusában, illetve a szuverénnek való alávetett-

ségben nyilvánul meg, aki viszont felette áll minden szerződéses viszonynak, őt nem kötelezi azok szerződése. Maga a szerződés olyan teremtő aktus, amelynek révén megalkotják azt a testet, amely egyszerre tartozik hozzájuk és idegenül el tőlük, így a test több, mint az őt alkotó alakok összessége. Csakis ez képes kifejezni azt az egységet, ami a szuverén intézményének, illetve eszközének, az állami erőszakmonopóliumnak mint a természeti állapot felszámolásának előfeltétele. A szuverén az őt alkotók védelmét, békéjét és biztonságát szolgálja, ugyanakkor az alakok mint pikkelypáncél borítják a szuverén testét, és ezzel ők maguk védelmezik azt. Mindez abban a cirkuláris érvelésben ölt formát, amelynek kiindulópontját a természeti állapot háborús viszonyai jelentik. Ezeket az a szerződés számolja fel, amelynek során megeremtik a Leviatánt, hogy a természeti állapot ne állhasson elő újra. Ezért felruházzák annak az erőszaknak a jogával, amelynek mindenki aláveti magát, hogy életét és biztonságát szavatolja. Az alattvalók maguk védelmezik a szuverént, hogy az megvédhesse őket egymástól és önmaguktól. Azért adnak korlátlan erőszakot a kezébe, hogy az erőszakot megfékezhesse. A Leviatán egyszerre véd és fenyeget. Egyszerre tartja fenn racionális számítás és irracionális félelem. Ebben rejlik az emberi alak alapvető kettőssége, amely csak ezen az önmagába visszatérő érvelés révén válik egységgé, jóllehet elsődlegesen mindig a képviselő és nem a képviseltek egységéről van szó.

Az emberek sokasága *egyetlen* személyben testesül meg, ha egy ember vagy személy képviseli őket, feltéve, hogy ez a sokaság minden egyes tagjának külön-külön hozzájárulásával történik. Mert a képviselő, és nem a képviseltek egysége teszi a személyt egyetlen személlyé. És a személyt a képviselő testesíti meg, de csak egyetlen személyt képvisel – a sokaságban egységet másképpen nem képzelhetünk el (Hobbes 1999. 200).

A test metaforája azonban nemcsak az egységet és az erőt fejezheti ki, hanem a szuverén hanyatlását és halandóságát is. Hobbes szerint az állam vagy fogantatásától kezdve beteg, „mint mikor beteg szülőktől származó gyermekeknek vagy idő előtt el kell pusztulniuk, vagy a rossz fogamzásból eredő káros tulajdonságaiktól a testükön keletkezett daganatok és gennyedések felmetszése útján kell megszabadulniuk” (Hobbes 1999. 330), vagy megbetegszik a szabadság elirigylésétől, a tirannofóbiától, a hatalommegosztás vágyától vagy a korrupciótól, amelyeket Hobbes a viszketegség, víziszony, eskór (nyavalyatörés, epilepszia) és mellhártyagyulladás metaforáival azonosít (Hobbes 1999. 334). Akik lázító gondolataikkal mételyeznek az államot, és „állandóan beleártják magukat a törvényhozás dolgába”, azokat egyszerűen galandférgeknek nevezi (Hobbes 1999. 338–339). Az alattvalók szervezeteit az államhoz való viszonyuk alapján az emberi testrészekhez hasonlítja, amennyiben a törvényesek az izmok, a törvénytelenek pedig a kelevények, daganatok és gennyedések (Hobbes 1999. 263). Ezt

a patológikus metaforikát némileg ellensúlyozzák a kereskedelem és mozgás, illetve a pénzforgalom és vérkeringés képei, hiszen „a mesterséges ember ebben is hasonlít a természeteshez, mert annak erei a test különböző részeiből kapott vért a szívbe vezetik, ahol is a vért a szív elevenné teszi, és az ütőereken át újból kiküldi, hogy ily módon a test valamennyi tagját elevenné és mozgóképessé tegye” (Hobbes 1999. 274).

Ezekből a metaforákból is kitűnik, hogy Hobbes mindent az egység elvének rendel alá. Jóllehet ebben az esetben még nem egy differenciált államról van szó, és retorikája is inkább a testben szerveződő organikus államfelfogásra emlékeztet, már megteszi a döntő lépést az új állameszme felé.

Mert a művészet teremtette meg a NÉPKÖZÖSSÉGNEK (latinul: CIVITASNAK) nevezett nagy LEVIATÁNT, ez pedig nem egyéb, mint mesterséges ember, bár természetes mintaképénél – amelynek védelmére és oltalmazására szánták – jóval nagyobb méretű és erejű; s benne a szuverenitás a mesterséges lélek, amely az egész testet élettel tölti meg és mozgatja; a tisztviselők, valamint a törvényhozó és végrehajtó hatalom többi képviselői a mesterséges ízületek; a jutalom és büntetés (amelyek a szuverenitás székhelyéhez fűződve az ízületeket és végtagokat feladatuk végrehajtására serkentik) az idegek, ugyanazt végzik, mint a természetes testben; ereje – valamennyi tagjának jóléte és gazdagsága; feladata – a *salus populi* (a nép üdve); tanácsadói, akik tájékoztatják minden számára szükséges tudnivalóról – az emlékezőképesség; a méltányosság és a törvény – a mesterséges értelem és akarat; az egyetértés – az egészség; a lázadás – a betegség, a polgárháború – a halál. És végül azok a szerződéses és egyezményes, amelyek ennek a politikai szervezetnek egyes részeit létrehozták, összeillesztették és egységbe foglalták, Istennek a teremtés alkalmával elhangzott *fiat* vagy *Teremtsünk embert* szavaihoz hasonlíthatók (Hobbes 1999. I. 69–70).

Mivel a természetesnek tekintett állapot a mindenki mindenki elleni harca, ezért az egyesülést nem elegendő kijelenteni, vagy a természetből levezetni. Szükség van egy mesterséges alkotásra, amely mindenkit egy testben egyesít, és ösztönszerűségét felülírva ésszerűségét helyezi előtérbe. Mindez egy alapítási akтуhoz kötődik, amely a szuverén személyében képes megteremteni a sokak egységét. Ez azonban nem a *koinónia politiké* vagy a *societas civilis*, hanem egy anyagtalan test, amely a politikai közösségtől tulajdonképpen független. A Leviatán testét alkotó, egymástól elkülönülő egyedek egyesülése nem a politikai közösség egységét fejezi ki, hanem a szuverénét. Talán éppen ezzel magyarázható, jóllehet zavarba ejtő, hogy a Leviatán alakjában szinte semmi félelmetes nincs; tekintete nem kelt félelmet, legfeljebb méreteivel és jelvényeivel fenyeget. Mindez különösen azért figyelemre méltó, mert Hobbes számára éppen a



fenyegetés, illetve a félelem a Leviatán meghatározó vonása, itt azonban az alak sokkal inkább kelt biztonságérzetet, mintsem halálfélelmet. Márpedig a Leviatánnak egyszerre kell elrettentenie és bizalmat ébresztenie. Mindez azonban már átvezet a Leviatán metaforájának azon részéhez, amely sem a képből, sem a szövegből nem olvasható ki, legfeljebb a bibliai utalásból fejthető meg.

#### 4. A szörny

Joggal merülhet fel a kérdés, hogy Hobbes vajon miért adott ennek az antropomorf-mechanikus konstrukciónak vallási-mitológiai nevet. Vajon miért a Leviatánt választotta a modern állam megtestesítőjének? A válasz túlmutat a kép és szöveg által közvetített hármasság egységen, ami az isteni, az emberi és a gépi között – ha nem is feszültségmentesen, de – fennáll. A bibliai Leviatán alakja kiegészíti ezt a hármasság képet, és mint szimbólum éppen azt a mozzanatot jeleníti meg, amely az eddigiekben nincs jelen. A Leviatán azzal nyerte el végső metaforikus sűrűségét, hogy Hobbes paradigmátikus szembeállította a Behemóttal – *Jób könyvének* másik szörnyével –, s ennek révén politikafilozófiai hagyományt teremtett.<sup>3</sup> Hobbes utolsó politikafilozófiai műve, a *Behemót, avagy a hosszú parlament* párbeszéd az angol polgárháborúról, amelyet 1670 körül írt, de a mű csak jó tíz évvel később, 1682-ben kapott nyomtatási engedélyt. Eszerint a *Jób könyvének* szörnyei arra hivatottak, hogy Jób Isten mindenhatóságát illető kétségeit eloszlassák: „semmivé teheted-é te az én igazságomat; kárhoztathatsz-é te engem azért, hogy te igaz légy? És van-é ugyanolyan karod, mint az Istennek, mennydörgő hangon szólalsz-é, mint ő?” (Jób 40, 3–4). A Behemót „az Isten alkotásainak remeke (és) az ő teremtője adta meg neki fegyverét” (Jób 40, 14). A Leviatán „lenéz minden nagy állatot, ő a király a gőg fia felett” (Jób 41, 34). A látomásokból születő hagyomány szerint a Leviatán tengeri lény, amely általában tengeri sárkányként vagy cethalként jelenik meg, míg a Behemót szárazföldi állat, amelyet legtöbbször elefántként vagy vízilóként ábrázolnak. Eredetileg a két szörny ugyanabban az összefüggésben tűnik fel, legfeljebb az általuk kép-

<sup>3</sup> A 20. században sajátos formában élt tovább a Leviatán és a Behemót kettős hagyománya. Carl Schmitt *Der Leviathan in der Staatslehre des Thomas Hobbes* című művében fenntartja ugyan Hobbes megkülönböztetését és az erős állam eszméjét, de állam- és alkotmányelméleti érvelését a mitológiaival elegyítve Hobbes szemére veti, hogy liberális, vagy legalábbis szabadságjogokat lehetővé tevő szférája miatt nem képes az egységet megteremteni – mindezt azzal az összeesküvés-elmélettel alátámasztva, amely szerint a zsidók minden egység akadályai. A Leviatán és a Behemót szembeállításának hagyományát folytatja Franz Neumann a *Behemoth – Struktur und Praxis des Nationalsozialismus 1933–1944* című, a nemzetiszocializmusról szóló elemzésével. Ebben a Leviatán a szuverenitás révén megteremtett erőszakmonopólium birtokosa, míg a Behemót az uralmi szempontból strukturált anarchia és polgárháború. A Behemót mint a náciizmus metaforája itt egyrészt a szuverenitás és a szabadság lerombolója, másrészt a káosz teremtője.

viselt princípiumokban – női–férfi, tengeri–szárazföldi – különböznek, jóllehet ebben a tekintetben is inkább kiegészítik egymást. A középkori keresztény teológiában a Behemót és a Leviatán a sátán különböző megtestesülései, de úgy tűnik, hogy Hobbes egy másik hagyományt követett, s ezt az értelmezés kapcsán nem hagyhatjuk figyelmen kívül. Ez a nézet arra a zsidó tradícióra megy vissza, amely a Leviatán nevét a héber *liv'iah*-ból vezeti le, amelynek eredeti jelentése: „összeilleszt”, „egyesít” (Bredekamp 2009. 25). Ez a vonás megszabadítja a Leviatánt gonosz jellegétől, és ennek megfelelően a későbbi értelmezésekben az uralkodó és az alattvalók közösségének egységeként jeleníti meg, ami a szuverén testében egyesülő sokaságra utal (Bredekamp 2009. 26). Ezzel szemben a Behemót alakja arra az egyiptomi hagyományra megy vissza, amely szerint a víziló az ösztönszerűség és a káosz jelképe. Hórusz isten Széthet vízilóként ejti el, ahogyan a rituális vízilóvadászat a rendet veszélyeztető erők feletti győzelmet szimbolizálta (Herz 1996. 263). Ez az eredendő hagyománybeli ellentét, amelyet Hobbes követ, életművében metaforává kristályosodik. Mindkét szörny Isten teremtménye és mindenhatóságának bizonyítéka, illetve mindkettő a hatalomra utal, de politikai szándékaik és kiindulópontjaik alapjaikban különböznek. Itt már nem egymás komplementereként, hanem antagonisztikus ellentétként jelennek meg. Leviatán a szuverenitás egységének és a békének, Behemót a szuverenitás felbomlásának, a polgárháborúnak és a természeti állapotba való visszahanyatlásnak a metaforája. Leviatán az uralom és a rend, Behemót a bomlás és a háború. Hobbes ezeket paradigmatikusan szembeállítja. Ha a kiindulópontot a Behemót felől nézzük, akkor világossá válik, hogy a Hobbes által leírt természeti állapot maga a polgárháború, amelyben nincsenek korlátok és szabályok, ezért mindennél súlyosabb pusztításhoz és végül az államrend felbomlásához vezet. Hobbes számára ez a legfőbb rossz, amit a szó szoros értelmében minden áron meg kell akadályozni. Az erőszakot csak erőszakkal, a szörnyet csak szörnyel, a Behemótot csak a Leviatánnal lehet leküzdeni.

## II. A LEVIATÁN RENDJE

A nyugati gondolkodás egyik sajátossága, hogy gyakran dichotómiákban, szimbolikus antagonizmusokban gondolkodik, és ezek többnyire időbeli sorrendet feltételeznek az egyes állapotok között. Így áll a helyzet a rend és a káosz kettős-ségével is. Hobbes feltételezett természeti állapota, amely kora polgárháborús viszonyainak felel meg, olyan viszonyokra utal, amelyek között az erőszak önkényes és kiszámíthatatlan, ezért ez magának a káosznak a szinonimája. Ezzel szemben a Leviatán fentebb leírt valamennyi alakjának háttérében a rend fogalma áll, a Leviatánban a rend eszméje és eszközei elválaszthatatlanul összefonódnak, illetve intézményesülnek (vö. Anter 2003). A rend mint eszme, illetve az állam és erőszakmonopóliuma mint eszköz kölcsönösen feltételezik egymást,

és csakis akkor beszélhetünk az egyikről vagy a másikról, ha összekapcsolódnak a szuverén gondolatában, illetve metaforájában, a Leviatánban. A modern állam kezdettől fogva elválaszthatatlan a rend eszméjétől, amelynek egyik oldala a félelem, a másik a biztonság, de ezek tulajdonképpen ugyanazon érme különböző oldalai (vö. Anter 2003). Ebben az összefüggésben csak a Leviatán képes arra, hogy a vallásháború félelmetes harcának véget vessen, és megteremtse az emberi élet legelemibb feltételét: a békét. Ehhez a rendhez szorosan kapcsolódik az uralmi igény, amely így egyedülálló igazolásra és létjogosultságra tesz szert. A modern állam szinte magától értetődően feltételezi a rendet, amely azonban így már nem egyszerűen állapot, hanem olyan kritérium, amely az uralmat, illetve az annak való elköteleződést legitimálja. Ez a kettős indoklás arra az egységre utal vissza, amely szerint a rend a Leviatán által jelképezett állam oka és következménye egyben.

Hobbes gondolkodásának újdonsága azonban nem egyszerűen ennek felismerésében áll, hiszen minden uralom rá van utalva a rend valamilyen formájára. De amíg a premodern korban a szuverén magát az isteni rend képének vagy egyenesen földi megfelelésének tekintette, addig a modern kori rend egyre inkább elszemélytelenedett és bürokratikusán szervezetté vált. A rend organikus szemlélete egyre inkább eltolódott a politika és társadalom mechanikus felfogásának az irányába. A modern kort megelőzően a rend elsősorban a szemlélés és megértés, illetve a rend fenntartásának problémája volt. Ha ugyanis a rend az isteni teremtés lenyomata, akkor az ember magát is azon rend szerves részének tekinti, amely örök és romolhatatlan. Ennek következménye az uralom tekintetében, hogy az a kozmoszról szóló elképzeléseknek felel meg, és ha mindkettőnek Isten a teremtője, akkor az uralommal kapcsolatban fel sem merülhet az igazságosság vagy igazságtalanság kérdése. E kérdés csak akkor vetődik fel, amikor a Leviatánban megtestesülő állam már nem egy isteni eredetű metafizikai intézmény, hanem az ember belső gyengeségéből adódó szükségszerűség (Bredenkamp 2001. 19). A modern kortól kezdve a rend egyre inkább a teremtés problémájává vált: immáron nem örök és nem isteni adomány, hanem történelmileg változó és fragmentált jelenség, illetve emberi tevékenység eredménye (vö. Anter 2005).

Ennek nyomán a rend fogalma egyre inkább dinamizálódott, és a történelmi folyamatba ágyazódva haladássá vált, ami arra utal, hogy a rendet meg lehet valósítani, csak személytelen törvényeit kell felismerni és szükségszerűséggé tenni. Hobbes éppen azt a pillanatot ragadja meg a Leviatán testében, amikor ez a folyamat végbemegy, és a rend végleg az emberi konvenciók kérdése lesz; amikor a rend már nem örök és eleve adott, hanem emberi eredetéből adódóan törékeny és állandóan újrateremtendő lesz. Hobbes bármennyire törekedett is a Leviatán révén az emberi állapot stabilitásának megteremtésére, a rend problémáját ezzel nem látta és nem is láthatta egyszer s mindenkorra megoldottnak vagy megoldhatónak (vö. Anter 2005). Így született meg az az új rendfogalom, amelynek alapján kénytelenek vagyunk tudomásul venni, hogy a rend képes nem az állapot örök, hanem maga a feladat.

## IRODALOM

- Anter, Andreas 2003. Im Schatten des Leviathan – Staatlichkeit als Ordnungsidee und Ordnungsinstrument. In Petra Bendel – Aurel Croissant – Friedbert W. Rüb (szerk.) *Demokratie und Staatlichkeit: Systemwechsel zwischen Staatsreform und Staatskollaps*. Opladen, Leske–Budrich. 35–55.
- Anter, Andreas 2005. Die Idee der Ordnung – Zur Paradoxie einer Leitfigur des politischen Denkens. In Andrea Gawrich – Hans J. Lietzmann (szerk.) *Politik und Geschichte – „Gute Politik“ und ihre Zeit. Wilhelm Bleek zum 65. Geburtstag*. Münster, Westfälisches Dampfboot. 14–29.
- Balke, Friedrich – Leander Scholz 2007. Die Disziplinierung des Bildes, Imagination und politische Ordnung. *Transkriptionen. Newsletter des Kulturwissenschaftlichen Forschungskollegs „Medien und kulturelle Kommunikation“ SFB/FK 427*. 8. 35–36.
- Bredenkamp, Horst 2001. Ikonographie des Staates: Der Leviathan und seine neuesten Folgen. *Leviathan. Berliner Zeitschrift für Sozialwissenschaft*. 29/1. 18–35.
- Bredenkamp, Horst 2003. *Thomas Hobbes, Der Leviathan. Das Urbild des modernen Staates und seine Gegenbilder. 1651–2001*. Berlin, Akademie Verlag.
- Bredenkamp, Horst 2009. *Behemoth* als Partner und Feind des Leviathan. Zur politischen Ikonologie eines Monstrums. *TranState Working Papers*. 98.
- Därmann, Iris 2006. Die Maske des Staates. Zum Begriff der Person und zur Theorie des Bildes in Thomas Hobbes' Leviathan. In Mihran Dabag – Kristin Platt (szerk.) *Die Machbarkeit der Welt*. München, Wilhelm Fink Verlag. 72–92.
- Herz, Dietmar 1996. Bürgerkrieg und politische Ordnung in Leviathan und Behemoth – Zum Kapitel 29 des Leviathan. In Wolfgang Kersting (szerk.) *Thomas Hobbes. Leviathan oder Stoff, Form und Gewalt eines bürgerlichen und kirchlichen Staates*. Berlin, Akademie Verlag. 259–281.
- Hobbes, Thomas 1999. *Leviatán, vagy az egyházi és világi állam formája és hatalma*. Ford. Vámosi Pál. Budapest, Kossuth Kiadó. I–II. kötet.
- Ludassy Mária 2007. Halandó Istenség és mesterséges örökkévalóság: A „Leviatán” metaforái. *Holmi*. 19/1. 45–53.
- Maye, Harun 2007. Die imaginäre Gemeinschaft. *Transkriptionen. Newsletter des Kulturwissenschaftlichen Forschungskollegs „Medien und kulturelle Kommunikation“ SFB/FK 427*. 8. 15–18.
- Voegelin, Eric 1938/1996. *Die politischen Religionen*. München, Wilhelm Fink Verlag.

## A szimpátia és a belénkszületettség testi értelmezése Descartes-nál

Ritka az olyan olvasó, aki ne érezné magát némiképp zavarban, amikor a *Hatodik elmélkedés* olvasása során hirtelen azzal szembesül, hogy életszellemekről és kötelekhez hasonlított idegekről szerez ismereteket. Legalább ennyire megmosolyogtatónak tűnnek az *Értekezés a módszerről* ötödik részében olvasható sorok az újbor forrásának és a szív mozgásának az összehasonlíthatóságáról. Értetlenségünket persze bizonyos mértékben csillapíthatja az a tudat, hogy a fenti szövegek láthatóan mechanikai elvekre hivatkoznak az idegrendszer és a vérkeringés leírása során. Ezt első ránézésre azonban nehezen tudjuk feltételezni például a *Dioptrique* alábbi részletéről:

Többek között azt is megmutathatnám önöknek, hogyan képes [az agy belső felületére nyomott kép] olykor egy terhes nő artériáin keresztül a magában hordott gyermek valamely meghatározott testrészt jutni, és ott megformálni az anyajegyeket, amelyek akkora csodálatot váltanak ki minden tudósban (AT VI. 129).<sup>1</sup>

Descartes szűkszavúsága az anyajegyek keletkezését illetően többek között azért aggasztó, mert meglehetősen nagy magabiztossággal társul: a szerző, a *Traité de l'homme* egyik szöveghelyéhez hasonlóan,<sup>2</sup> itt is úgy lép fel, mint aki világosan ismeri a jelenség megfelelő magyarázatát, ám nem kíván eltérni az adott értekezés szigorúan vett tárgyától. Egy Mersenne-nek írott levélben ugyanakkor fellelhetünk egy olyan megközelítést, amely önmagában még nem felel meg a magyarázat Descartes által adott kritériumainak:<sup>3</sup> ezek szerint például Panni testén azért találhatóak ilyen és ilyen gyümölcsök alakját idéző anyajegyek, mert Panni édesanyja a terhessége során leginkább ilyen és ilyen gyümölcsöket

A tanulmány a 81165 számú OTKA-pályázat támogatásával készült.

<sup>1</sup> Az AT rövidítés Descartes 1996-ra utal.

<sup>2</sup> Vö. AT XI. 177 (a szöveget a későbbiekben idézem).

<sup>3</sup> AT III. 120. Descartes még azt is hozzáfűzi itt, hogy a gyermek kigyógyítható az anyajegyektől, ha a megfelelő gyümölcsöt fogyasztja el.

kívánt volna fogyasztani. A *lélek szenvedélyei* második részében olvasható példa nyomán pedig beláthatjuk, hogy az anyai aktivitás nem csupán külső testfelületünk sajátosságait befolyásolhatja; ezek szerint egyesek a rózsa illatával szembeni eleve adott averzióval jönnek világra, mégpedig azért, mert édesanyjuk is erős averzióval rendelkezett a rózsailleslattel szemben terhessége során (vö. AT XI. 429. 136. cikkely).

A fenti és a hozzájuk hasonló szöveghelyek a *Traité de l'homme* első, 1664-es francia kiadását kommentáló Louis de la Forge aggodalma szerint azért szorulnak minél pontosabb magyarázatra, mert a descartes-i fiziológia leghomályosabb pontját alkotják;<sup>4</sup> ugyanis közös sajátjuk, hogy valamilyen fizikai összefüggést tételeznek fel az anya agyában megjelenő „képek” és az embrió fejlődésének karakterisztikumai között. Az alábbiakban arra teszek kísérletet, hogy Descartes fiziológiai írásai alapján rekonstruáljam az összefüggés azon mechanisztikus magyarázatát, amelyre a filozófus egyes töredékeiben csak implicit módon utal. Első lépésben a testi érzékelés és az emlékezet descartes-i leírását vizsgálom majd.<sup>5</sup> Ezután, elsősorban a *Description du corps humain* negyedik és ötödik fejezete alapján, felvázolom az embrió fejlődésének állomásait, hogy ezekből kiindulva értelmezem a „szimpátia” Descartes által vázolt példáit. Végezetül olyan következtetések megfogalmazására kerül sor, amelyek a descartes-i filozófia átfogóbb kérdéseit érintik.

A következtetések során érintett problémák összetettsége jól mutatja, hogy bár a fiziológiai írások sokak szemében a descartes-i életmű marginális jelentőségű darabjainak tűnnek, e szövegek elemzésének nem csupán a tudománytörténeti anekdoták szintjén van tétje. Az alább értelmezendő szövegek hozzájárulhatnak a Descartes-ról alkotott képünk további pontosításához: azokat a problémákat figyelhetjük meg ugyanis, amelyek a „mechanikai filozófia” gyakorlatba ültetésének során merülnek meg fel. Emellett a testi folyamatok descartes-i leírásának adekvát megközelítése segítségünkre lehet a „dualizmus” címszóval jelölt kérdéscsoport tisztázásában is: az állati lelkek létezésének problémájával kapcsolatos descartes-i álláspont korántsem egyszerűsíthető le azon a módon, ahogyan azt hinni szoktuk; s a fiziológiai írások elemzése az emberi lélek és az emberi test „összeillesztettségének” mibenlétéről is komplexebb értelmezést nyújthat.

<sup>4</sup> Vö. La Forge – Descartes 1664. 335. De la Forge úgy fogalmaz, hogy e pontok homályban hagyása a teljes descartes-i vállalkozás értékét fenyegetné.

<sup>5</sup> Azért nem a „testi képzelet” kifejezést használom, mert bár Descartes a később idézett szöveghelyeken a képzelet működését kívánja értelmezni, nem teljesen egyértelmű, hogy akkor is így tenne-e, ha nem feltételezné folytonosan a *Traité de l'homme* ezen szakaszában, hogy a testhez lélek is társul. Az aktív és a passzív képzelet elkülönítése ugyanis nem feltétlenül korrelál az intellektuális és a testi képzelet elkülönítésével. Igaz, az „érzékelés” kifejezés is némiképp megrévesztő lehet a *L'homme* kontextusában: az éhség érzését ugyanis Descartes a lélekhez köti. Az alábbiakban a szót a „testi képek agyi felületekre történő nyomódása” kifejezés rövidítéseként használom.

## I. AZ IDEA TESTISÉGE: AZ ÉRZÉKELÉS ÉS AZ EMLÉKEZET

Ahhoz, hogy megérthessük, hogyan nyomódnak az anya agyában előforduló képek az embrió testére, először is azt kell megértenünk, hogy hogyan kerülnek egyáltalán képek az anya agyába. A *Traité de l'homme* szövegének utolsó negyede arra próbál magyarázatot adni, hogy miként működik a testi érzékelés és az emlékezet, és hogy hogyan bírják az érzékelés és az emlékezet képei mozgásra a testet. A mű leírásai szerint egy tárgy látványának testi ideájára a következő lépésekben teszünk szert.<sup>6</sup>

Először is, a tárgy felületéről visszaverődő fény különböző módokon nyomást gyakorol a szemfenékre, pontosabban a szemfenéken található bizonyos idegvégződésekre (vö. AT XI. 173). Ezáltal a szemfenéken egy olyan alakzat (*figure*) jelenik meg, amely hasonlít (Descartes a *se rapporter* igét használja) a tárgyhoz.<sup>7</sup>

Másodsor: ezen alakzat az idegfonalak közvetítésével az agy belső felületére jut (vö. AT XI. 175). Az idegek Descartes leírásában olyan csövek, amelyek felülete valójában az agyat körülölelő bőr meghosszabbítása, és amelyekben egyrészt az érzékszervektől az agy felületéig futó fonalak, másrészt pedig, e fonalak körül, az életszellemek harmadik, a vér legfinomabb részéből álló típusának<sup>8</sup> tömege található.<sup>9</sup> A szellemek az érzékelésnek ezen kezdeti szakaszán egyedül azt szolgálják, hogy rugalmassá tegyék az idegek fonalait: ezek, a szellemektől körülvéve, anélkül képesek elhajlani, hogy elszakadnának. Az agy belső felületén megjelenő alakzat ugyanakkor Descartes leírása szerint még nem tekinthető ideának.<sup>10</sup>

Az idea szerepére ugyanis még testi értelemben is egyedül a tobozmirigy felületén megjelenő nyom tarthat számot (vö. AT XI. 176). Az érzékelés harmadik

<sup>6</sup> A következő lépések leírását lásd AT XI. 170–179.

<sup>7</sup> Vö. AT XI. 175. 12–13. sor. Descartes azért nem az *image*, és azért inkább a *figure* szót használja, mert leírása szándékai szerint nem csupán a látásra nyújt magyarázatot, de minden más szerv által történő érzékelésünkre is. Ehhez lásd AT XI. 176.

<sup>8</sup> Egy 1643-ban Vorstiusnak írt levelében Descartes a testben lévő „szellemek” három típusát különítette el: a „természetes szellemek” (*spiritus naturales*) az emésztés során keletkeznek, és sajátosságaikat döntően meghatározzák a táplálékok; az „életszellemek” (*spiritus vitales*) akkor keletkeznek, amikor a szív felhevíti a vért, és belőle kiválasztja a finomabb részeket; ezek közül pedig azok, amelyek a carotison át egyenes vonalban haladva képesek az agyba lépni, nem mások, mint az „animális szellemek” (*spiritus animales*). A szellemek minden fajtájáról igaz, hogy megtévesztő nevük ellenére csak és kizárólag testi jelleggel bírnak (a *spiritus* kifejezés itt inkább a „légiességre” utal; Descartes maga is a levegőhöz hasonlítja különösen a *spiritus animalest*). Vö. AT III. 687. A magyar fordítói hagyomány sajnos szándékolatlanul is egyenlőségjelet tett az „életszellem” és a szellemek harmadik típusa közé.

<sup>9</sup> Az idegek descartes-i anatómiáját lásd AT XI. 132–133. Descartes e pontokon lényegében Galénossal vitatkozik, aki a *De usu partium* VII. könyvének 5. fejezetében két típusú ideget különít el: az egyik kemény, és a mozgásért felelős, a másik lágy, és az érzékelést szolgálja. Ezzel szemben Descartes egyetlen idegtípus létezését ismeri el, és az idegeken belül futó fonalakat teszi felelőssé az érzékelésbeli képek megjelenítéséért, míg az ezek körül áramló szellemek mozgatják az izmokat. Vö. Galien 1859. I. 539–541.

<sup>10</sup> Az alábbiakban az „idea” kifejezést abban a testi értelemben használom, amelyet a *Traité de l'homme* is követ (vö. AT XI. 177).

lépésének megértéséhez először is azt kell tudnunk, hogy a tobozmirigy az agy legbelső pontján helyezkedik el,<sup>11</sup> és a belőle kiáradó életszellemek veszik körül. Ezen életszellemek kiáradásának mértékét és irányát egyebek között az agy belső felületén megjelenő kép határozza meg. Amikor az alakzat e felület bizonyos pontjaira érkezik, e pontokon kitágítja az idegi csatornák kimenetét; az így meghatározott kimeneteken pedig az életszellemek szabadon áramolhatnak át – ennek következtében pedig a tobozmirigy azon pontjain távoznak elsősorban, amelyek korrelálnak a belső felületen megjelenő alakzat pontjaival. E korreláció természetét Descartes nem határozza meg pontosan, és csupán annyit jegyez meg, hogy a szellemek a tobozmirigy egy adott pontján inkább hajlanak arra, hogy egy adott irányba mozogjanak, mint egy másikba. Ha az agy belső felületén olyan út nyílik meg, amely egy adott ponton elhelyezkedő, és így egy adott irányú mozgási hajlandósággal rendelkező szellemek hajlandóságának kedvez, akkor az ezen ponton lévő szellemek haladnak át a tobozmirigy felületén (vö. AT XI. 175–176). E felületen ezáltal a szellemek kiáradása maga is nyomot hagy, és így az alakzat a tobozmirigy belső felületén is megjelenik, és miután a descartes-i magyarázat szerint ehhez fér „a legközvetlenebb módon hozzá” a lélek, ezt az alakzatot tekintjük ideának (vö. AT XI. 177).

Megjegyzendő, hogy a tobozmirigy kitüntetettsége a legszélesebb körben vitatott descartes-i elképzelés; látnunk kell azonban, hogy ezen a ponton nem annyira a lélekkel való kapcsolata, mint inkább a testi képek összefogásában betöltött szerepe a fontos. Descartes ugyanis abból indul ki, hogy például a két szemideg nem találkozik egymással, tehát elvileg két különféle képet visz az agy felé; a tobozmirigy mint nem páros szerv pedig azért felelős, hogy e két képből egy álljon elő.<sup>12</sup> Az így megformálódó képek ugyanakkor nem csupán az érzékelés, de az emlékezet alapjául is szolgálnak: ha egy tárgyat gyakran érzékelünk, akkor a tobozmirigy adott pontjain kiáradó életszellemek jobban kitágítják az agy belső felületén található idegi kimeneteket (amelyeket a fonalak által továbbított alakzat nyomott a belső felületre), és így a tárgyat akkor is láthatjuk, ha nincsen jelen (vö. AT XI. 177–178). Ráadásul a képek képesek előhívni egymást: mivel a tűz látványáért felelős agyfelületi kimenetek általában együtt nyílnak meg a melegség érzetéért felelős kimenetekkel, a tűz látásával együtt felidéződik bennünk a melegség emléke is (vö. AT XI. 179).

Különös ugyanakkor, és témánk szempontjából kulcsfontosságú, hogy Descartes a következő bekezdést ékeli az érzékelés és az emlékezet testi folyamatának leírása közé a *Traité de l'homme*-ban:

<sup>11</sup> Descartes a *Traité de l'homme*-ban a tobozmirigyre „H-mirigyként” utal. Elhelyezkedéséhez lásd AT XI. 174.

<sup>12</sup> Ezt a magyarázatot adja például *A lélek szenvedélyei* 32. cikkelye is a tobozmirigy kitüntettségére. Vö. AT XI. 352–353.



Hozzáfűzhetném még mindehhez, hogy hogyan jutnak ezen ideák nyomai az artériákon keresztül a szív felé, és sugároznak így szét a vér egészében; és hogy hogyan képesek ezek olykor az anya bizonyos cselekvései (*actions*) által arra kényszeríttetni, hogy belenyomódjanak a hasában formálódó gyermek testrészeibe. Itt azonban megelégszem azzal, hogy elmondom önöknek, hogyan nyomódnak rá az agy belső részére [...], ahol az *emlékezet* székhelye található (AT XI. 176).

Feltevésém szerint annak lehetősége, hogy a „nyomokat” az artériák a test felé továbbíthatják, Descartes számára azért idézendő fel ezen a ponton, mert e lehetőség nélkül nem adhat teljes egészében számot arról, hogy hogyan mozog a külső testek hatására, de a lélek jelenléte és/vagy beavatkozása nélkül a test.

Vegyük például a *Traité de l'homme* egyik legtöbbször idézett leírását. Ezek szerint, ha kezünk közel kerül a tűzhöz, akkor a következőképpen reagálunk a lélek közreműködése nélkül is: egyrészt elrántjuk onnan a végtagunkat, másrészt a fájdalom forrása felé fordulunk (vö. AT XI. 142). A folyamat a fentiek alapján és némi kiegészítéssel részben értelmezhető is: a tűzről leváló részecskék nagy erővel esapódnak a kéz felületén lévő idegvégződésekhez; az idegek mozgása kitéríti az agy belső felületének bizonyos pontjait, amelyen át az életszellemek nagy mennyiségben a kezet vezérlő izmokhoz áradnak. Ezen izmok azért tanúsítanak mozgást, mert az életszellemek, pontosan nagy mennyiségük okán, egy légballoonhoz hasonlóan felfújják azokat (pontosabban az opponáló izmok egyikét),<sup>13</sup> és ezáltal mozgást keltenek a kézben. Mindez azonban csak és kizárólag a kézízmok reakciójára ad magyarázatot; az egyáltalán nem derül ki, hogy mi okozza a többi testi mozgást, így például azt, hogy fejünket a tűz felé fordítjuk; a fejünket mozgó idegek tájkáról ugyanis semmilyen hatás sem érte a külvilág részéről a fejünket.<sup>14</sup> Úgy tűnik tehát, hogy olyan komplex testi reakcióval van dolgunk, amelyet nem magyarázhatunk egyedül az érzékelés fent vázolt lefolyásával; és, ha nem akarunk a lélek közreműködésére hivatkozni, akkor fel kell tennünk, hogy bizonyos idegi kapcsolódások már embrionális korban létrejöttek. Ennek megértéséhez fel kell vázolnunk a magzati fejlődés descartes-i leírakísérletét.

<sup>13</sup> Az izmok felfúvódásával kapcsolatban lásd AT XI. 134.

<sup>14</sup> Descartes itt úgy fogalmaz, hogy a kézízmokban végződő fonalak megrándulásának hatására az agy belső felületén lévő pórusok megnyílnak, és e pórusokon keresztül a szellemek „egy része” a kéz felé, más „része” a fej felé stb. veszi az irányt (vö. AT XI. 143. 10. sor). A probléma azonban az, hogy e szellemek e „megoszlását” egyáltalában nem motiválja semmi az érzékelési folyamat fent idézett leírásában, ráadásul ellentmondani látszik az idegek descartes-i anatómiájának.

## II. A MAGZATI FEJLŐDÉS ÁLLOMÁSAI

Descartes az 1630-as évek kezdetétől fogva többször kísérletet tett a magzati fejlődés lépéseinek kijelölésére.<sup>15</sup> Az alábbiakban elsősorban az 1638 és 1648 között írt *Description du corps humain* leírásaira támaszkodom majd, de bizonyos pontokon figyelembe veszem a korábbi töredékeket is. Meg kell jegyeznünk ugyanakkor, hogy a különböző időpontokban keletkezett töredékek jelentős mértékben ellentmondanak egymásnak, mégpedig leginkább azon kérdést illetően, hogy milyen sorrendben alakulnak ki az embrióban az egyes szervek; például az 1630-as évek elején keletkezett szövegek szerint először nem a szív, hanem az agy fejlődik ki a magzatban. Ezen ellentmondások ellenére azonban állandó elemnek tekinthetjük, hogy Descartes az embrió kifejlődésére is egyrészről teleológiai hivatkozások nélkül, másrészről pedig a *Monde*-ban és a *Principiá*-ban leírt mozgástörvények alapján próbál magyarázatot adni. A magyarázatok között tapasztalható eltérés pedig vélhetően abból következik, hogy Descartes a különböző időkben keletkezett szövegekben más és más, a kifejlett emberi testben leírt folyamatot választ azon analógia alapjának, amelynek mentén az embrió kialakulását értelmezheti.

Ez az analógiás alap a *Description* esetében a szív működése és a vér keringése. Descartes itt úgy véli, hogy a két nem egy-egy egyedéből származó folyadék egy olyan keveréket alkot (Descartes ezt nevezi itt magnak), amelyben a keveredő folyadékok egymással szemben kovászként lépnek fel; a kovászolódást pedig itt úgy kell elképzelnünk (vö. AT XI. 253), hogy a magfolyadék kitágul, részei ugyanis elválnak egymástól (bizonyos részei inkább hajlamosak erre, mint mások). Ez a folyamatleírás egyértelműen ekvivalens a szív működésnek a Plempius-levezetés idejétől (1637–1638) datálható descartes-i értelmezésével:<sup>16</sup> itt a szívben fellelhető, az előző kitágulásnál nem távozott maradványvér tágítja ki a szívbe érkező új vért annyira, hogy az a szív újabb kitágulásához vezessen.<sup>17</sup>

<sup>15</sup> Aucante nyomán az embriológiai írások három csoportját különíthetjük el: az első 1630 és 1632 között keletkezett; a második hozzávetőlegesen egy időben az *Értekezéssel* és az *Esszéikkel*; a harmadik az 1638 és 1640 között készülő *Description du corps humain* negyedik és ötödik fejezetével azonos. Az első két csoport szövegeit a *Cogitationes circa generationem animalium* (első kiadása: 1701) és az *Excerpta anatomica* (első kiadása: 1859) című töredékgyűjteményekben lelhetjük fel (a *Description*hoz hasonlóan ezek is az életműkiadás XI. kötetében találhatóak meg). Vö. Descartes 2000. 2–6.

<sup>16</sup> Plempiusnak az *Értekezés* ötödik fejezetével kapcsolatos ellenvetései hatására Descartes olyan magyarázati elemekkel igyekezett menteni a szív működés általa adott leírását, amelyek a *Traité de l'homme*-ban még nem szerepeltek; ilyen elem a reziduális vér kovászoló hatása.

<sup>17</sup> A *Description*-ban a maradványvér maga is kovászként kerül jellemzésre: „Az a kevés megritkított vér pedig, amely a szív kamráiban maradt, azonnal összekeveredvén azzal, ami újfólag kerül a szívbe, úgy viselkedik, mint egyfajta kovász, és azt okozza, hogy a vér hirtelen felhevül és kitágul; e módon pedig a szív bizonyos mértékben felfújódik, megkeményedik és megrövidül” (AT XI. 231).

A fejlődés első, keveredései szakasza után kezdetét veszi a belső szervek, a keringési rendszer és az idegek egyfajta folyékony előképének a kialakulása.<sup>18</sup> A *Description* szerint a mag dilatációs mozgása során kialakul egy olyan közép-pont, a szív kezdeménye, amely a tágulás koncentrációs, kiindulási pontjául szolgál; innen bizonyos részek, mivel természettől fogva egyenes vonalú mozgásra törekszenek, a szívkezdemény „ellenállását” legyőzve kilépnek, és létrehozzák az agyalapot; a helyükre lépő részek pedig körmozgást végeznek, amíg ki nem tágulnak maguk is (vö. AT XI. 254). Ebből is látható, hogy milyen tényezők bírnak kitüntetett szereppel az embrionális fejlődés descartes-i magyarázatában: a vákuum létének tagadása okán a mag részei alapvetően körmozgást végeznek, egyesek pedig bizonyos ideig képesek megvalósítani az egyenes vonalú mozgásra irányuló törekvéseiket. A mag részei által e mozgástörvények alapján bejárt utak pedig Descartes magyarázata szerint a később megszilárduló belső szervek egyfajta folyékony vázát formálják meg.

Miután ugyanis az előbb még a szívkezdeményben körmozgást végző részek maguk is elválnak egymástól, és egyenes vonalú mozgásba kezdenek, a szívből kilépve előbb létrehozzák az aorta felső ágát, az agyalapról visszaverődve pedig a gerincoszlopot, ahonnan lefelé haladva az aorta alsó ágának kezdeményét formálják meg (vö. AT XI. 257). Egy újabb dilatáció eredményeként az ezen leszálló ágon végigfutó vér szintén ellenállásba ütközik, és visszafordulván az üres vénát kezdi kialakítani. Az üres vénán felfelé haladó vér a szívhez érve kitágul, és létrehozza a szív második, jobb kamrájának előképét (vö. AT XI. 258). A következő „pulzálásnál” a jobb kamrából kiváló részek a tüdőartéria, a tüdő és a tüdővéna, a bal kamrából kiváló részek pedig az agy kezdeményeit formálják meg; ezen a ponton pedig megkezdődik a vérrészek differenciálódása: azok a részek, amelyek képesek a tüdőn át a bal kamrába, és onnan az agyba jutni, alkotják majd az életszellemeket (AT XI. 258). A vér ezen részei az agyban vagy ezen keresztül megkezdik az érzékszervek kialakítását (vö. AT XI. 261–264); a többi pedig a vénás és artériás rendszer kialakulásáért lesz felelős (vö. AT XI. 259). E folyékony fejlődési szakasz a köldökzsinór kialakulásával zárul (vö. AT XI. 273): az anyai vér innentől kezdve vesz részt az embrió fejlődésében.

A fejlődés e harmadik szakaszáról tudjuk meg a legkevesebbet (mind a *Description* szövegében, mind a töredékekében). Egyedül annyi tűnik bizonyosnak, hogy az embrió megszilárdulásának folyamata Descartes szerint párhuzamba áll

<sup>18</sup> Descartes már egy 1637-re datálható töredékében (AT XI. 516) is három szakaszát különíti el az embrionális fejlődésnek: az elsőben a mag felhevül, a másodikban kialakulnak a belső szervek, a harmadikban pedig már az anyai vér táplálja a magzatot. E háromsztatúság a *Description*-ban is tettenérhető, azonban a második szakasz leírása itt egyértelművé teszi, hogy ennek során a belső szerveknek csupán egyfajta folyékony előképe alakul ki. A magzati fejlődés három szakaszra való bontását Aucante érvelése nyomán Galénosznak tulajdoníthatjuk (vö. Aucante 2006. 301), aki a *Természetes fakultásokról* szóló írásában maga is három részt különít el a kifejlődési folyamatban (*generatio, auctio, nutritio*), igaz, teljességgel más kontextusban (vö. Galien 1859. 2. köt. 217).

lítható azzal,<sup>19</sup> ahogyan a kifejlett emberi test a vérből magához veszi a táplálékokat.<sup>20</sup> Ilyen szempontból különösen releváns az a magyarázat, amit Descartes a hízásról ad: ilyenkor a szerveket felépítő fonalak közé olyan mennyiségű folyadék lép be az erek pórusain keresztül, hogy az intervallumok túltelítetté válnak, és a fonalak mozgása lelassul (vö. AT XI. 249). Az embrió fejlődése a harmadik szakaszban Descartes számára így nem más, mint radikális hízás: a folyékony szervezdemények az anyai vérből származó részekkel telítődve előbb fonal-szerkezetűvé válnak, majd a fonalak közötti tér is részben feltöltődik általuk.<sup>21</sup>

### III. A SZIMPÁTIA REJTÉLYE

A fentiek alapján azt kell mondanunk, hogy a descartes-i fiziológia de la Forge által leghomályosabbnak minősített része ahhoz a fejlődési szakaszhoz csatlakozik, amely ráadásul a legkevésbé kifejtett, tehát a megszilárdulási folyamatokhoz; az ezt megelőző szakaszok ugyanis az anyai vértől függetlenül is lezajlanak. E homályosság és kifejtetlenség ellenére a „szimpátia” a *Traité de l’homme* említett szöveghelyei alapján mégis kulcsjelentőségűnek tűnik,<sup>22</sup> mert Descartes vélhetően komoly magyarázó szerepet tulajdonított az anyai agyműködés és a magzati fejlődés összefüggésének: részben ennek kell feltárnia a test kívüllagra adott reakcióinak alapját, de vélhetően annak magyarázatát is, hogy miért hasonlítanak a gyermekek szüleikre.<sup>23</sup> Mindenesetre Descartes explicit módon a szimpátia következő funkcióit említi embriológiai töredékeiben: először is (a) a szimpátia rendelkezik egy olyan jelentéssel, amely nem az embrióhoz, hanem a kifejlett testre vonatkozik: Descartes szerint egy lantművész művészetét vezérlő képek

<sup>19</sup> Valójában e párhuzam meglétéről is „fordított irányban” szerezhethünk értesülést, ugyanis a *Description* a kifejlett ember táplálkozási folyamatának leírása után, és az embrionális élet jellemzése előtt így fogalmaz (AT XI. 252): „Még tökéletesebb tudásra tehetünk szert arról, hogy hogyan tápláltanak a test egyes részei, ha megfontoljuk, hogy hogyan jönnek létre a magból”.

<sup>20</sup> E párhuzam részletezését lásd Gaukroger 2000. 389–391.

<sup>21</sup> Látunk kell, hogy Descartes-nál a „megszilárdulás” nem jelent úgymond „minőségi” változást. A „folyékony” és a „szilárd” testi részek elválasztásának alapját a *Description*-ban az adja, hogy a folyékonyabb testi részek összetevői gyorsabban, míg a szilárdabb testi részek összetevői lassabban mozognak (vö. AT XI. 247). Mindez összhangban áll a folyékony és a szilárdság *Monde*-ban bevezetett fogalmával (vö. AT XI. 10–16).

<sup>22</sup> Az anya és a magzat testének „szimpátiájára” való hivatkozás Descartes korában igen gyakori volt. A kortárs álláspontok és a történeti előzmények összefoglalásához lásd Aucante 2006. 122–125. Ezen árnyalt elemzéseket leegyszerűsítve azt mondhatjuk: a szimpátia a kortárs fiziológiai diskurzusban egymástól távol eső testrészek olyan együttes elváltozása (*compassio*), amelynek során a változást elsődlegesen és közvetlenül magába fogadó rész a saját elváltozását valamilyen módon a másik részre is átírja. E szimpátia alapesete az emberi test egészének „harmonikus működése”; kitüntetett és konkrét megvalósulása pedig az anyai és a magzati testet köti össze.

<sup>23</sup> A szülő és a gyermek hasonlóságának, illetve a szimpátia kapcsolatának részletes kifejtését lásd Smith 2006. 89–95.

nemcsak az agy belső felületét határozzák meg, de onnan szétsugározva a test más részeit is (például a kezét).<sup>24</sup> Másodszer (b) ritkán ugyan, de Descartes valamilyenfajta egybehangoltságot tételez fel a magzat elhelyezkedése, szituáltsága és az anyai test között, sőt esetenként az anya és a magzat bizonyos testrészeinek megfeleltethetőségét állítja.<sup>25</sup> Harmadszor (c) úgy véli, hogy az anyai agyi aktivitás befolyásolhatja a kifejlődő embrió testrészeinek alakját (az anyajegyek esete); végezetül (d) több helyütt azt állítja, hogy az anyai idegi kapcsolatok megjelenhetnek az embrió agyában is (a belénkszületett averziók esete). Az alábbiakban elsősorban a (c) és a (d) jelű esetek vizsgálatára kerül sor, feltevésem szerint ugyanis ezek alapján egyrészt az (a) és a (b) könnyen értelmezhetővé válik, másrészt pedig világosan mutatják majd a szimpátiára vonatkozó leírások térjének kiterjeszhetőségét. Látnunk kell ugyanakkor, hogy a (c) és a (d) esetek lehetőségfeltételét az (a) eset biztosítja; azonban az előbbi két eset lefolyása sokkal könnyebben megragadható Descartes írásai alapján, így ezek alapján fogok következtetni az (a) eset magyarázatára.

Az ezen esetek megközelítését célzó szöveghelyek azonban első látásra egyáltalán nem illeszthetőek könnyedén a testi gép descartes-i leírásai közé. Ha a *Traité de l'homme*-ból származó fenti idézetet vesszük alapul, akkor a nyilvánvaló problémát az okozza, hogy (1) egyáltalán nem világos, hogy hogyan „sugározhatnak szét” a tobozmirigyből *közvetlenül* az artériákon keresztül a *szív felé* a „képek”; ehhez ugyanis vagy azt kell feltételeznünk, hogy a szellemek képesek visszafelé áramlani az artériákban (ami Descartes más megfogalmazásai alapján képtelenségnek tűnik), vagy azt, hogy Descartes vénákat akart említeni artériák helyett. Ez utóbbi lehetőséget azonban (nyilvánvaló problematikussága mellett) kizárja, hogy a fiziológiai írások minden esetben csak a tobozmirigy artériáira, de sohasem a vénáira hivatkoznak.<sup>26</sup> (2) Szintén problémát jelent, hogy Descartes nem igazán részletezi annak módját, ahogyan a szellemek harmadik típusa visszakerülhetne a véráramba: az világos ugyan, hogy a vér legfinomabb részei képesek belépni az agyba, az azonban teljesen tisztázatlan marad, hogy hogyan is léphetnének ki onnan, és hogy hogyan keringhetnek a véredények rendszerének egészében. (3) Feltéve, hogy az utóbbi problémát kizárjuk, az továbbra sem tisztázható, hogyan őrződhet meg a tobozmirigy belső felületén megjelenő

<sup>24</sup> A lantjátékos példájának kifejtését egy Mersenne-nek szóló levélben olvashatjuk (vö. AT III. 48). Ugyan Descartes az emlékezet működésének megvilágítása érdekében hivatkozik a példára, a lantjátékos esetét mégis a szimpátiához sorolhatjuk, mert – ami mechanizmusokat illeti – a kettő megegyezik egymással.

<sup>25</sup> Erre például szintén egy Mersenne-levélben olvashatunk utalást (vö. AT III. 120–121), amely szerint „általában véve a gyermek egyes testrészei megfelelnek az anya egyes testrészeinek”. Descartes ezt egy olyan asszony példájával szemlélteti, aki terhessége alatt eltörte a kezét, és gyermeke is törött kézzel jött világra.

<sup>26</sup> A *Traité de l'homme* szerint a tobozmirigy az artériák függesztik fel az agyban (vö. AT XI. 179).

„nyom” az életszellemekben; e „nyom” ugyanis nem a szellemek, hanem a szellemek és a tobozmirigy interakciójának az eredménye.

A La Forge által javasolt megoldás (La Forge – Descartes 1664. 336–338) az említett közül csak az első problémát oldja meg (részlegesen). Ezek szerint a tobozmirigyből az artériába lépő életszellemek megváltoztatják az eleve az artériában lévő életszellemek mozgási sebességét; ez a változás azonban „visszafelé” is hatást gyakorol a szellemek viselkedésére, ugyanis az artérián „lejjebb lévő” életszellemeknek is fel kell gyorsulniuk ahhoz, hogy a „feljebb lévő” életszellemek helyét kitölthessék, illetőleg le kell lassulniuk, mivel nem találnak helyet a tovahaladásra.

Bár La Forge válasza nem meggyőző megoldás, sokkal jobbat mégsem javasolhatunk. Egyedül az vihet bennünket – igaz, legfeljebb egy lépéssel – közelebb a megoldáshoz, ha megfontoljuk: Descartes leírása szerint az agy minden a vérképzéssel és a vér minőségével kapcsolatba hozható szervvel idegi összeköttetésben áll, tehát képes befolyásolni azt is, hogy milyen jellegű vér érkezzon az embrió testéhez. Tegyük fel például, hogy Panni anyja mindennél jobban körtét kívánna enni; ez többek között azt jelenti, hogy egy körte láttán a szívmozgása megváltozik (pontosan azon a tisztázatlan módon, ahogyan a testtartásunkat is megváltoztatjuk a tűz kezünkhöz való közelsége okán), ugyanis a szívideg úgy kezdi befolyásolni a billentyűket, hogy azok nagyobb mennyiségű vért erresszenek át egyetlen dilatáció alkalmával is.<sup>27</sup> Tegyük fel továbbá, hogy a körte látványa áttételesen ugyanígy befolyásolja a máj működését. Mindez azt jelenti, hogy az aortába, majd onnan a placentába és a köldökzsinórba jutó vér, illetőleg az első és a második típushoz sorolható életszellemek egészen más minőségeket fognak mutatni, „rendes esetben”; ez a minőségváltozás pedig befolyásolja az embrió bőrének kialakulását, és felelőssé lesz az anyajegyek kialakulásáért. Ugyanígy adhatunk számat a rózsza iránti belénkszületett averzió létezéséről, csakhogy itt az embrióba lépő anyai vérnek nem az a része a számottevő, amely a bőr vagy más szervek megszilárdulásában, hanem amelyik az agy belső, „emlékezeti” felületének a megformálásában vesz részt (tehát az, amely képes az embrióban életszellemmé alakulni).

<sup>27</sup> A *Traité de l'homme*-ban a szívideg egyfelől a szenvedélyelmélet kontextusában jelenik meg (AT XI. 164–165): „[H]a a szívbe érkező vér tisztább, finomabb, és könnyebben gyullad lángra (*embrase*), mint rendes esetben, akkor azon a módon rendezi el az ott található apró ideget, amely az *öröm* érzetének okozásához szükséges; ha pedig e vér teljességgel ellentétes minőségekkel bír, akkor az ideget a *szomorúság* érzetének okozásához szükséges módon diszponálja”. Másfelől azonban Descartes azt is leszögezi, hogy a szívideg döntően befolyásolja a szellemek képződését: „Végezetül pedig mindaz, ami valamely változást képes okozni a vérben, ugyanúgy változást képes okozni a szellemekben [...] Mindenekfelett azonban az az apró ideg, amely éppúgy képes kitágítani és összehúzni azt a két bejáratot, amelyen át a vénák vére és a tüdő levegője a szívbe ereszkedik, mint azt a két kijáratot, amelyen át a vér kiárad és kitör az artériákba, képes ezernyiféle változást előidézni a szellemek természetében” (AT XI. 169).

E megoldással némiképp megkerüljük a (2)-es számú problémát, azonban megoldásunk legfeljebb La Forge interpretációjával együtt lehet érvényes, és annak feltételezésével, hogy valóban létezik egy ennél közvetlenebb módja annak, hogy a szellemek a véráramba kerüljenek (hacsak nem akarjuk teljesen metaforikus értelemben venni Descartes idézett sorait). A (3)-as számú probléma élet némiképp elvethetjük, ha felidézzük Descartes alábbi, Mersenne-nek címzett sorait az anyajegyek problémájával kapcsolatban:

Ami az anyajegyeket illeti, önt csak az bírja rá arra, hogy úgy higgye: e jegyek tökéletesen hasonlítanak a tárgyakra, hogy különösnek találja azt, hogy ezek egyáltalán hasonlítanak a tárgyakra; ha azonban összeveti őket a legsilányabb módon megfestett képekkel, úgy találja majd, hogy e jegyek még azoknál is tökéletlenebbek (AT III. 49).

Descartes úgy véli tehát, hogy az anyajegyek valóban hasonlíthatnak bizonyos tárgyakra, azonban a hasonlóság mértéke olyan kicsiny, hogy még egy rajzból bukott diák festménye is jobban emlékeztet egy körtére, mint egy körte alakú anyajegy. Mindez azt jelenti, hogy az életszellemek csak meglehetősen kis sikerrel képesek a véren át az embrió testéhez szállítani a tobozmirigy felületén megjelenő alakzatokat. Vélhetően ezt hivatott ellensúlyozni a szimpátia folyamatának túldeterminált jellege (az [1]-es és a [2]-es problémák megoldására javasolt folyamatok egymástól függetlenül, de egymáshoz adódva befolyásolják a vér jellegét), valamint, hogy Descartes sok esetben „erőtlen” képzeleti képeket említ az anyai oldalon, amikor a szimpátia kérdését érinti.

E sok eset ugyanakkor nem fedi az összeset. A szimpátia (b)-vel jelzett esete vagy jelentése alapvetően nem a kiemelkedő erőt megtestesítő képeken nyugszik, hanem feltehetőleg a vér és az életszellemek rendes áramlásán. Descartes úgy véli, hogy az anyai szervek megfeleltethetőek (itt franciául a *correspondre*, latinul a *respondere* igét használja)<sup>28</sup> az embrionális szerveknek (de nem minden esetben közvetlenül); e megfeleltetés pedig talán azzal magyarázható, hogy az a vér és azok a szellemek, amelyek az anyában leginkább egy adott szervet táplálnak, az embrióban is azt táplálják majd, vagy egy konkrétan meghatározott és kijelölt másik, megfeleltetésben álló testrészt táplálnak. Mindenesetre a vér és a szellemek itt is ugyanazon, megnyugtató módon nem tisztázott mechanizmusok folytán áramlanak az anya testéből az embrióba.

Összefoglalva azt mondhatjuk, hogy a szimpátia mechanizmusának pontos leírása nem rekonstruálható egyértelműen Descartes írásai alapján; legfeljebb fiziológiájának azon elemeit tudjuk kiemelni, amelyek lehetővé teszik, vagy elősegítik az anyai agyi aktivitás és az embrió fejlődése közötti kapcsolat fennállását. A kérdés ekkor az, hogy meggyőző mechanikai magyarázat híján miért

<sup>28</sup> Vö. AT III. 121 és AT XI. 533.

ragaszkodik Descartes a legkorábbi írásoktól egészen a legkésőbbi szövegekig a szimpátia meglétének hangsúlyozásához. Descartes ugyanis még az alábbi, legkésőbbre datálható embriológiai töredékében is a szimpátiával foglalkozik:

Valamiféle szétsugárzásként (*radiationem*) kell értelmeznünk azt, ami a magzatban reprodukálja az anya által nagyobb figyelemmel elgondolt dolgokat; fel kell tennünk, hogy a magzat úgy helyezkedik el az anyaméhben, hogy a feje az anya feje felé áll, a háta az anya háta felé, a jobb oldala pedig az anya jobb oldala felé; a vér pedig az anya fejéből egyenlően oszlik el a magzat körül, hogy aztán a köldökszínóiban gyűljön össze, ahonnan ezen összegyűlés ellentétéként útját a magzat minden része felé folytatja (AT XI. 538).

E szöveg jól mutatja mindazt, ami kiderült, és mindazt, ami nem derült ki a fentiek alapján: az anyai agyműködés, sőt, esetlegesen az anyai gondolkodás és a magzat teste között az anya vére valamilyenfajta kapcsolatot teremt; ráadásul az anya testrészei és az embrió testrészei között eleve korreláció áll fenn. Az azonban ebből a töredékből sem válik világosabbá, hogy e folyamat miként zajlik le; a kérdés tehát az, hogy mi szüksége Descartes-nak a szimpátiára,<sup>29</sup> ha nem tud rá magyarázatot adni.<sup>30</sup>

#### IV. KÖVETKEZTETÉSEK

A fentiek alapján lehetőségünk nyílik olyan következtetések levonására és olyan sejtések megfogalmazására az említett kérdés kapcsán, amelyek messze túlmutatnak a descartes-i fiziológia tárgyalásán, és a hagyományosan többet tárgyalt filozófiai problémák értelmezéséhez járulhatnak hozzá. Az első ilyen probléma az állati „lélek” létével vagy nem létével kapcsolatos. A második az emberi lélek természetét érinti: véleményem szerint a szimpátia fogalma némiképp segíthet tisztázni azt, hogy mit értett Descartes a lélek egész testben való szétterjedésén. Végezetül pedig a szimpátia esetleges etikai jelentőségét tárgyalom.

<sup>29</sup> Des Chene jegyzi meg többször, hogy Descartes fiziológiája meglehetősen „konzervatív” természetű abban az értelemben, hogy minél kevesebb magyarázóelv segítségével törekszik minél több jelenség magyarázatára (vö. Des Chene 2001. 50). A fentiek azt mutatják, hogy jelenségek egy egész sora marad magyarázat nélkül, ha nem ismerjük el a szimpátia létezését.

<sup>30</sup> Egy kézenfekvőnek tűnő megoldás persze felmerül, különösen, ha Malebranche felől olvassuk Descartes-ot: Descartes a szimpátia fogalmának alkalmazásával kívánta elkerülni azt, hogy például a szörnyszülöttek létezéséért Istent kárhoztassuk. Azonban miközben igaz, hogy Descartes megemlíti a szimpátia kapcsán a szörnyszülötteket (vö. AT XI. 518), Malebranche későbbi tézise nem jelenik meg nála explicit módon. A szörnyszülöttekkel kapcsolatos malebranche-i gondolatokhoz lásd Schmal 2006. 44–46.



1. Fontenelle a *Lettres Galantes* egyik darabjában így ír:<sup>31</sup>

Azt állítja, az állatok és a szörnyek csupán gépek? Helyezzen hát egymás mellé egy kan kutyagépet és egy szuka kutyagépet, és minden bizonytalán egy harmadik apró gépet kap eredményül; ám két órától, még ha egész életükben is egymás mellett ülnek, sohasem lesz egy harmadik. Márpedig mi filozófiánkban [...] úgy találtuk, hogy mindazok a dolgok, amelyek ketten vannak és rendelkeznek annak képességével, hogy egy harmadikat hozzanak létre, sokkal emelkedettebbek és nemesebbek, mint a gépek (Fontenelle 1742. 99).

E tréfás összehasonlítás megfogalmazásával Fontenelle a sokak szerint egyébként is ezer sebből vérző descartes-i fiziológia alapvetése ellen intéz támadást: Fontenelle álláspontja szerint a jelenségek egy döntő részére nem tudunk magyarázatot adni akkor, ha a testet pusztán gépezetnek tekintjük. A problémát így fogalmazhatjuk meg: ha egyszer, Descartes állításának megfelelően, az állatok nem rendelkeznek lélekkel, amelynek számára az idegek és az életszellemek mozgása „alkalmat adhatna” arra, hogy például éhséget vagy szomjúságot érzzenek, akkor egyáltalán mi indíthatja őket arra, hogy egyenek vagy igyanak; és hogyan magyarázható, hogy ezen indítatásokkal már születésüktől fogva rendelkeznek?

E probléma kapcsán először is látnunk kell, hogy Descartes-nál az állati lélek létezésének tagadása korántsem olyan egyértelmű, mint ahogy azt hinni szoktuk. Ugyan kétségtelen tény, hogy az *Élmélkedések* megírása idején Descartes tagadja, hogy az állatok rendelkeznek lélekkel, mondván, hogy nem gondolkodnak; az *Értekezés* és az *Essais* körül kialakult viták egyik alapvető dokumentumában, a Fromondusnak írt levélben azonban például így fogalmaz:

Világosan elmagyaráztam, hogy az állatok lelke semmi más, mint a vér, amely, a szívben szellemekké melegedvén és finomodván, az artériáktól az agyig, és onnan az idegekig és az izmokig terjed (AT I. 414).

Mindazt mutatja, hogy az 1630-as évek Descartes-ja szerint az állatok ugyan nyilvánvaló módon nem rendelkeznek racionális lélekkel, de bizonyos (később egyébként Descartes által zavarónak minősített) értelemben mégis beszélhetünk az állatok lelkéről – ha elfogadjuk, hogy e lélek anyagi természetű. Mindaz tehát annyit jelent, hogy a vérként felfogott „lélek” alapján kellene számot adnunk arról, hogyan értelmezhető például a szomjúság (és a rá következő ivás) jelensége például egy disznó esetében.

<sup>31</sup> A szöveg helyi jelentőségére Des Chene tanulmánya hívta fel a figyelmemet (vö. Des Chene 2001. 12).

A kulcsot mindehhez, úgy vélem, a szimpátia fogalma szolgáltatja: a disznók és más állatok vélhetően úgy jönnek a világra, hogy agyukban a (bennünk, emberekben) szomjúságérzésre alkalmas idegi mozgások már eleve össze vannak kapcsolva a szellemek olyan irányú mozgásával, amelynek folytán az állati test víz keresésére indul. A *Cogitationes circa generationem animalium* egyik töredékében ugyanis ezt olvashatjuk:

Az állatok semmilyen fogalommal sem rendelkeznek az előnyösről vagy a károsról, ám a méhben való létezésük során a növekedésükért felelős erő által olyan dolgoknak voltak kitéve, amelyek folytán bizonyos mozgásokra ösztönöztek (*impulsa sunt*); és így mindig, amikor valami hasonlóval találkoznak, ugyanezen mozgásokat hajtják végre (AT XI. 520).

Az anyai „lélek”, tehát a vér, amely a fentiek értelmében képes képeket és kép-kapcsolatokat szállítani a vénákon, az artériákon és az idegen át, ennek alapján olyan erőként kerülhet értelmezésre, amely az embrionális fejlődés mozgását befolyásolva befolyásolja azon „lélek” természetét is, amely a magzatban, majd pedig az újszülöttnél lehetséges fel.

2. A lélek szenvedélyei első részében egy igen zavaró szöveghelyre bukkanhatunk. A harmincadik cikkelyben Descartes így fogalmaz:

[T]udnunk kell, hogy a lélek valójában a teljes testtel össze van illesztve, és hogy nem állíthatjuk tulajdonképpen értelemben, hogy bármelyik részében helyezkedne el a többi kizárásával. Ennek oka egyrészt az, hogy a test egy, és bizonyos értelemben oszthatatlan, mégpedig szerveinek elrendeződése alapján, amelyek úgy kapcsolódnak egymáshoz, hogy amint az egyik közülük hiányzik, az egész test tökéletlenné (*défectueux*) lesz. A másik ok abban áll, hogy a lélek olyan természettel bír, amelynek semmilyen-fajta kapcsolata sincs a kiterjedt dolgokkal, sem a dimenziókkal, sem azon anyag bármely más tulajdonságaival, amelyből a test össze van téve; a lélek egyedül a testi szervek összeállításának (*assemblage*) egészével áll kapcsolatban (AT XI. 351).

Zavarunk oka teljesen világos: az *Elmélkedések* bizonyos szöveghelyei alapján azt kell mondanunk, hogy szigorú értelemben véve a léleknek semmilyen közössége sincs a testtel, mégpedig pontosan azért, mert a test osztható, míg a lélek oszthatatlan. Idézetünk ezzel szemben azt állítja, hogy e közösség létezik, mégpedig pontosan azért, mert a *test* bizonyos szempontból oszthatatlan; a lélek pedig maga is oszthatatlan lévén a test elrendezésének egészéhez (és nem a test egy kitüntetett részéhez) csatlakozik. Ráadásul közvetlenül ezután Descartes leszögezi, hogy a lélekkel való kapcsolódásának szempontjából testünknek mégiscsak van

egy kitüntetett pontja, nevezetesen a tobozmirigy; így nem is annyira az okoz problémát, hogy milyen kapcsolatban is áll a lélek a tobozmiriggyel, hanem hogy e két álláspont hogyan egyeztethető egyáltalán össze.

Úgy vélem, a szimpátia jelenségének figyelembevétele nélkül e descartes-i elmélet teljességgel érthetetlen marad. Térjünk most vissza a szimpátia eddig nem tárgyalt, (a)-val jelölt típusához. Ezek szerint egy lantművész nem csupán agyában, de kezében és teljes testében az, mert a játéka alapjául szolgáló képeket emlékezete nem csupán az agy belső felületén, hanem a véren keresztül a test más tájékain is elraktározza. Itt természetesen a testhez, és nem a lélekhez kötődő emlékezetéről van szó; de Descartes példája *A lélek szenvedélyei* szövegére vetítve értelmezhető úgy is, hogy ugyan a lélek legközvetlenebbül a tobozmirigy képeihez fér hozzá, kevésbé közvetett módon azonban a test többi részében raktározódó benyomásokkal is operálhat (Panni, ha lantművész, például jóval könnyebben bírja kezét lantjátékra, mint én). A lélek tehát azért terjed szét az egész testben, mert az észlelésre és a cselekvésre alkalmas adó képek maguk is szétterjednek a vér által. A fenti szöveghely alapján így azt mondhatjuk, hogy a testi mozgások és a lelki tevékenységek megfeleltetésének (amely megfeleltetés fennállását Descartes az *institution* szóval jelöli) nem csupán a tobozmirigy, hanem az egész test működése szolgáltat alapot.

3. Az előzőekben azt láttuk, hogy a szimpátia fogalmára való hivatkozás kulcsjelentőségű például az állatok lelkének létezését vagy az emberi lélek egész testben való szétterjedését érintő problémák tárgyalásakor. Befejezésül (de leginkább egy sejtés formájában) azt szeretném megmutatni, hogy nem csupán a szimpátia fogalmába foglalt folyamat hatókörének kiterjesztése, de egyben szigorú korlátozása is fontossággal bír – mégpedig a descartes-i etikai gondolkodás területén.

Ennek belátáshoz először is azt kell figyelembe vennünk, hogy a descartes-i természetfilozófiában előfordulnak olyan „aszimmetriák”, amelyek explicit módon etikai analógiákkal bírnak. Ezen „aszimmetriák” közül talán a legtöbbet tárgyalt az egyenes vonalú mozgás és a körmozgás viszonyát érinti. A *Monde* harmadik természeti törvényének jellemzése végén ezt olvashatjuk:

E szabályból következően azt kell mondanunk, hogy egyedül Isten a világban fellelhető összes olyan mozgás szerzője, amely egyenes vonalú; szabálytalanná és görbe vonalúvá pedig csakis az anyag különféle diszpozíciói teszik e mozgásokat. Ugyanígy tanítják számunkra a teológusok, hogy Isten mindazon cselekvésünk szerzője, amelyek léteznek, és amelyek tartalmaznak valamennyi jóságot; bűnössé pedig csakis akarásaink különféle diszpozíciói teszik e cselekvéseket (AT XI. 46–47).

Úgy tűnik, e szöveghely egyfelől „szabályossági” különbséget tételez fel az egyenes vonalú és a görbe vonalú mozgás között: az egyenes vonalú mozgás szabályos, mert közvetlenül Isten az oka; e szabályosságot pedig azon anyagi környezet ronthatja le, amelybe az adott mozgó test illeszkedik. Másfelől pedig idézetünk párhuzamot von e „szabályossági” eltérés és cselekvéseink értékülönbségei között: eredendően és Istentől adódóan a jóra hajlanánk, ám akarásaink eltérítenek ettől; teremtésünkéből adódóan etikailag „egyenszek” vagyunk, de akarásaink által „görbékké” válunk.

Hasonló, analógiás kiegészítéssel terhelt aszimmetriára figyelhetünk fel a melegség és a hidegség, és vele párhuzamban a szeretet és a gyűlölet között. E gazdag analógia csupán egyik elemét kiemelve azt láthatjuk,<sup>32</sup> hogy míg a szeretet jót tesz az egészségnek, addig a gyűlöletről mindez nem állítható; az pedig, hogy miért tesz jót a szeretet az egészségnek, részben azzal magyarázható, hogy megélénkíti a test melegét, míg a gyűlölethez éppenséggel hideg társul (vö. AT XI. 402. 97. cikkely). A szeretet és a gyűlölet aszimmetriája így egy sajátos fiziológiai szinten alapozódik meg: mivel a testben a melegség eredendőbb, mint a hidegség, hiszen e melegség mozgatja a szívet, és kelti életre az embriót, az etika szintjén a melegséggel társuló szeretet is eredendőbb szerephez jut, mint a gyűlölet (annak ellenére, hogy a descartes-i célkitűzések értelmében egyik affektus sem rendelkezhetne kitüntetettséggel a másikkal szemben).

Mit tudunk mindehhez hozzáfűzni a szimpátia fogalmának vizsgálata után? Itt válik fontossá, hogy Descartes szerint az embrionális fejlődés első két szakaszát nem befolyásolja az anyai vér, következésképpen az anyai agyi működés sem. Ennek megfelelően a magzat szíve önmagától kezd verni, és szervezdeményeinek többsége kialakul azelőtt, hogy az anyai vér tápláló erejéből részesülne. Ebből következően viszont azt kell mondanunk, hogy az anyai test, még ha a leghidegebb gyűlölettel átitatva él is, annyira nem tudja befolyásolni az embrió fejlődését, hogy az ne örizné meg eredendő melegségét. A szimpátia hatókörének korlátozottsága nem azért érdekes, mert nem tudja megváltoztatni a belső szervek felépülését (ami egyébként korántsem bizonyos), hanem mert nem tudja elvenni azt a magzati hőt, amely a szeretet fiziológiai korrelátuma. Máshonnan megközelítve a kérdést: bár a szimpátia működése odavezethet, hogy az anya egy szörnyszülöttet hoz a világra, sőt az is előfordulhat, hogy a véletlen folytán a magzat például nem nélkülivé válik, e szörnyszülött vagy sérült újszülött Descartes szerint ugyanazt az eredendő hőt, és így ugyanazt az „eredendő jósságot” hordozza önmagában, mint bármelyikünk.

Ezt megfontolván viszont részben választ adhatunk két, eddig fel nem vetett kérdésre: egyfelől arra, hogy Descartes miért úgy jelöli ki az embrionális fejlődés szakaszainak határait, ahogyan kijelöli, másfelől pedig arra, hogy végül miért a szív működését választja az embrió fejlődésének megértéséhez szükséges ana-

<sup>32</sup> Ezzel kapcsolatban lásd Boros 2003. 109–127.

lógia alapjául. Az első kérdésre azt válaszolhatjuk, hogy Descartes számára pontosan azért kulcsfontosságú a magzat „folyékony vázának” kialakulása az anyai vér beavatkozása előtt, hogy az „eredendő hő” „átöröklődése” semmiképpen se sérüljön, még akkor sem, ha extrém esetben még a fajspecifikus sajátosságok is távol maradnak az embriótól. Ennek kapcsán pedig a második kérdésre is válasz adódik: Descartes azért választja a szív működését az embrionális fejlődés modelljéül, mert így tudja biztosítani azt, hogy az „eredendő hő” mindenképpen megjelenjen (és meg is maradjon) az embrióban – hiszen a szív az, ami a kifejlett ember esetében is az egész testbe melegséget sugároz szét.

## IRODALOM

- Aucante, Vincent 2006. *La philosophie médicale de Descartes*. Paris, Presses Universitaires de France.
- Boros Gábor 2003. *A természettörvényektől Isten értelmi szeretetéig*. Budapest, Áron.
- Des Chene, Dennis 2001. *Spirits and Clocks: Machine and Organism in Descartes*. Ithaca, Cornell University Press.
- Descartes, René 1996. *Oeuvres*. Szerk. Charles Adam – Paul Tannery. Paris, Vrin. XI kötet (= AT).
- Descartes, René 2000. *Écrits physiologiques et médicaux*. Paris, Presses Universitaires de France.
- Fontenelle, Bernard de 1742. *Oeuvres de Fontenelle*. Paris. I–VIII. köt.
- Galien 1854–1856. *Oeuvres anatomiques, physiologiques et médicales de Galien*. Ford. Charles Darremberg. Paris, J-B Ballière. I–II. köt.
- Gaukroger, Stephen 2000. The Resources of a Mechanist Physiology and the Problem of Goal-Directed Processes. In Stephen Gaukroger – John Schuster – John Sutton (szerk.) *Descartes' Natural Philosophy*. London – New York, Routledge. 383–400.
- La Forge, Louis de – René Descartes 1664. *L'homme de René Descartes et un traité de la formation du foetus du même auteur, avec les remarques de Louis de la Forge*. Paris, Charles Angot.
- Schmal Dániel 2006. *Természettörvény és gondviselés: egy filozófiai és teológiai kérdés a korai felvilágosodás korában*. Budapest, L'Harmattan.
- Smith, Justin E. H. 2006. Imagination and the Problem of Heredity in Mechanist Embryology. In uő (szerk.) *The Problem of Animal Generation in Early Modern Philosophy*. Cambridge, Cambridge University Press.



Rovatunk ezúttal a magyar filozófiai szaksajtóban immár évek óta, jórészt e lap hasábjain zajló vita folytatásához kínál *fórumot*. A vitát kiváltó szöveg a *Kellék* című folyóirat analitikus filozófiai számának programadó cikke volt (Észes Boldizsár – Tózsér János 2005. Mi az analitikus filozófia? *Kellék*. 27–28. sz. 45–71), amelynek utolsó fejezete az analitikus filozófia és a kontinentális filozófiai irányzatok, különösen a fenomenológia viszonyát taglalta. A fenomenológusok reagálása nem késett sokáig (a legfontosabb írások: Schwendtner Tibor 2006. Analitikus versus kontinentális filozófia. *Magyar Filozófiai Szemle*. 50/1–2. 1–15; Ullmann Tamás 2006. Analitikus és kontinentális filozófia. *Magyar Filozófiai Szemle*. 50/3–4. 227–251); majd rövidesen megjelent egy közvetíteni kívánó írás (Faragó-Szabó István 2007. És hova álljanak a belgák? Hozzászólás az analitikus–kontinentális vitához. *Magyar Filozófiai Szemle*. 51/1–2. 109–127), amelynek azonban nem igazán volt mérséklő hatása az analitikus kezdeményezők viszontválaszának tartalmára (Észes Boldizsár – Tózsér János 2007. Mi és mi nem az analitikus filozófia? *Magyar Filozófiai Szemle*. 51/1–2. 75–108). Faragó-Szabó István személyében a vitatkozók mindenesetre megfelelő moderátorra találtak a vita szóbeli, az Eötvös Kollégiumban lebonyolított fordulójához. A nézőpontok ütközése olyan jellemzőnek tűnt filozófiai életünkre, hogy pozsonyi társlapunk magyar filozófiai tematikus számában is beszámoló jelent meg róla mint filozófiai életünk egyik legjellemzőbb eseményéről (Faragó-Szabó, István 2009. Spor medzi analytickou a kontinentalnou filozofiou. *Filozofia*. 64/2. 107–113).





## Analitikus filozófia és fenomenológia

Ullmann Tamás *Analitikus és kontinentális filozófia* (Ullmann 2006) című tanulmányában *Mi az analitikus filozófia?* (Eszes–Tözsér 2005) című cikkünk apropóján fejti ki álláspontját az analitikus filozófiáról és a fenomenológiáról, valamint azok viszonyáról. Tanulmányát három részre tagolja. Az első részben azt mutatja be, hogy szerinte melyek az analitikus filozófia legfontosabb, ugyanakkor tisztázatlan előfeltevései; a másodikban azt tárgyalja, hogy ezekről az előfeltevésekről miképpen vélekedik a fenomenológia; a harmadikban pedig egy lehetséges jövőképet vázol fel: hite szerint a két irányzat előbb vagy utóbb, valamilyen formában, de egyesülni fog.

Nagyon kevés dologban értünk egyet Ullmann-nal. Vitapartnerünk állításai az analitikus filozófia állítólag tisztázatlan előfeltevéseiről javarészt félreértéseken vagy tévedéseken alapulnak. Vitapartnerünk „fenomenológiai nézőpontból” (Ullmann 2006. 241) megfogalmazott állításainak megítélése már összetettebb feladat, összbenyomásunk azonban inkább kedvezőtlen. Úgy gondoljuk, Ullmann fejtegetései során számos fontos fogalom, fogalmi kapcsolat összemósodik, vagy torz megvilágításba kerül. Végezetül vitapartnerünk prófécijával rokonszenvezünk ugyan, de nem tartjuk sem különösebben életszerűnek, sem különösebben ésszerűnek azt, ahogyan a két, immár egyesült iskola jövőbeni munkamegosztását elképzeli.

Tanulmányunkban rámutatunk Ullmann főbb tévedéseire. Ennek során jól áttekinthető tanulmányának szerkezetét követjük. Nem vállalkozunk arra, hogy parafrázáljuk Ullmann gondolatait, álláspontját inkább idézetekkel adjuk vissza.

Tözsér Jánost e tanulmány megírásában az MTA-TKI támogatta, valamint a tanulmány írása során a Bolyai ösztöndíj támogatását élvezte. A szerzők köszönetüket fejezik ki Bárány Tibornak a tanulmányhoz fűzött remek észrevételeiért, és Márton Rózának, amiért kiegyengette a stílisis kacifántokat.

## I. AZ ANALITIKUS FILOZÓFIA ÁLLÍTÓLAGOS ELŐFELTEVÉSEI

Ullmann szerint az analitikus filozófia mögött „súlyos és észrevétlen előítélettömeg húzódik meg [...], amire az analitikus filozófia nem akar, vagy nem tud reflektálni” (Ullmann 2006. 230), és amely előfeltevések „súlyosan befolyásolják] a tapasztalatra, gondolkodásra, cselekvésre stb. irányuló leírásokat”. Ullmann öt ilyen „magától értetődően elfogadott” előfeltevést említ, amelyek közül szerinte „legalább egyet (de inkább többet)” valamennyi analitikus filozófus elfogad (Ullmann 2006. 231). Vegyük sorjában.

1. Ullmann a tanulmány *Fizikalizmus, naturalizmus, szcientizmus* című részében így fogalmaz:

Az analitikus filozófusok jó része, úgy tűnik, hallgatólagosan elfogad egy ontológiát, aminek a minimum elemei a következők: a világ különálló dolgok összessége, ezek a dolgok szubsztanciális létezéssel bírnak, vagyis egy változatlan hordozó mag körüli szervezett tulajdonságegyüttesként foghatók fel, a világban zajló események pedig e különálló dolgok közötti viszonyok. Arról, hogy mi van, s hogy az miképp van, leginkább a fizika tud minket felvilágosítani. Az analitikus filozófia végső hivatkozási alapja az újkori természettudomány. [...] Ez egyébként rendben is van, kevés izgalmasabb dolog van a fizikánál, a biológiánál, az agykutatásnál. Csakhogy ez nem jelenti azt, hogy a filozófiának el kell fogadnia ezek – éppen az eredményes kutatás miatt szükségszerű – módszertani korlátait. Úgy látom egyébként, hogy [...] az analitikus filozófia a természettudomány szédítően izgalmas újdonságait rendre visszarángatja a maga idejétmúlt tárgyontológiájába. Ennek a tárgyontológiának az az alapfeltevése, hogy a létezés alapformája a „tárgyi” létezés (mintha a biológiai önszerveződő rendszerek, az anyag-hullámok, az erők, a komplex neuronhálózatok, a rejtélyes természetű részecskék, a kaotikus rendszerek, a szuperhúrok, és még lehetne folytatni a sort, nem lennének a kortárs természettudomány érdeklődésének homlokterében) (Ullmann 2006. 232).

Nos, vegyük sorra Ullmann állításait. Kezdjük azzal, hogy valóban vannak olyan filozófusok (Allaire 1963/1998; 1965/1998; Bergman 1967; Martin 1980), akik John Locke nyomdokaiban haladva a fizikai tárgyakat „egy változatlan hordozó mag körüli szervezett tulajdonságegyüttesként fog[ják] fel” (Ullmann 2006. 232). Vagyis – precízebben fogalmazva – azt gondolják, hogy a fizikai tárgyak nem mások, mint tulajdonságok nélküli pusztán hordozók (szubsztrátumok) *plusz* különböző tulajdonságok együttese. Csakhogy e metafizikai elmélet *nem általánosan elfogadott* a kortárs analitikus filozófiában. (Hozzátennénk azt is, hogy sohasem volt az.) Számos analitikus filozófus élt és él (például Hochberg 1964,

Casullo 1984, 1988, Van Cleve 1985/1998), akik épp a szubsztrátum fogalmának (egyébként közismert) nehézségei miatt vélekednek úgy, hogy nem léteznek szubsztrátumok, és a fizikai tárgyak pusztán *univerzálék nyalábjai*. Valamint olyanok is szép számmal akadnak, akik – szintén tagadva a szubsztrátumok létezését – azt állítják, hogy a fizikai tárgyak pusztán *trópusok* (azaz partikuláris tulajdonságok) *nyalábjai* (például Williams 1953/2004, Campbell 1981/1998, 1990, Simons 1994/1998). A trópuselmélet híveit pedig meglehetősen méltánytalan azzal vádolni, hogy megfélekednek a „természettudomány szédítően izgalmas újdonságairól” (2006. 232), különösen annak fényében, hogy az elmélet egyik legfontosabb kortárs képviselője (Simons 1994/1998) igen behatóan vizsgálja a kérdést: hogyan egyeztethető össze a trópuselmélet a kortárs fizika által posztulált entitások (fermionok, bozonok stb.) létezésével. (A tulajdonságok metafizikájáról lásd részletesen Tözsér 2009.)

Folytatva kritikánkat: Ullmann tévesen állítja, hogy az analitikus filozófusok szerint az „események [...] különálló dolgok közötti viszonyok” lennének (Ullmann 2006. 232). Ullmann összekeveri az „esemény” fogalmát a „tény” fogalmával, ugyanis ténynek (és nem eseménynek) nevezzük – a partikuláris tárgyak tulajdonság-instanciálása mellett – a különböző partikuláris tárgyak között fennálló relációkat. Például tény (és nem esemény), hogy az asztalom nagyobb a széknél. Ugyancsak tény (és nem esemény), hogy a Parlament épülete közelebb van a Halászbástyához, mint a Hősök teréhez.

Ullmann akkor is téved, amikor minden megszorítás nélkül azt állítja, hogy az analitikus filozófia szerint „a létezés alapformája a »tárgyi« létezés” (Ullmann 2006. 232). E kijelentés azért nem igaz, mert számos kortárs analitikus filozófus nem tesz éles metafizikai különbséget a fizikai *tárgyak* és az *események* között (Quine 1960. 36. rész; 1963; 1976; 1981; Lewis 1986; 2001; 2002; Armstrong 1980; Heller 1990; 1993; Sider 1996; 2001). E filozófusok úgy tekintik a fizikai tárgyakat, mint amelyeknek – az eseményekhez *hasonlóan* – vannak időbeli (temporális) részei is. Például egy 20 centiméter hosszú drótnak nemcsak térbeli részei vannak (mondjuk a bal 10 centimétere és a jobb 10 centimétere), hanem időbeli részei is: mint például „a drót tegnap”, „a drót ma”, „a drót holnap”. Ezt az irányzatot perdurantizmusnak nevezzük (részletesen lásd Tözsér 2009).

Még valami. Ullmann szerint az analitikus filozófia állítólagos tárgyontológiájának súlyos – az analitikus filozófusok számára „nyilvánvalóan” átláthatatlan – következményei vannak. Idézzük:

Ha minden létező tárgyi létező, akkor az alapviszonyok mind kauzális viszonyok: ezért a természetleírás nem más, mint kauzális leírások összessége. Ebből következnek azután az olyan furcsa gondolati képződmények, hogy ha létezik elme, akkor annak okoznia kell. Így jutunk el az „elme-kauzalitáshoz” (hogyan okoz az elme testi mozgást?), ami nem egy kérdés

pontos körülhatárolását, hanem a zavart és a hagyományos [...], de a valóságos problémáktól már messzire eltávolodott fogalmiság tovább erőltetését jelenti (Ullmann 2006. 233).

Szerintünk e passzus egyetlen helyes következtetést sem tartalmaz:

- (a) Abból, hogy minden létező tárgyi létező, *nem következik az*, hogy az alapviszonyok mind kauzális viszonyok, mert az oksági viszonyoknak egészen bizonyosan nem a fizikai tárgyak a relátumai, azaz az oksági viszony nem fizikai tárgyak között fennálló reláció. A legtöbb kortárs filozófus szerint események között áll fenn oksági viszony.
- (b) Abból, hogy léteznek oksági viszonyok a világban, *nem következik az*, hogy a természetleírás kauzális leírások összessége, mert plauzibilisen lehet érvelni amellelt is, hogy a fizika törvényei (a természettörvények) nem oksági természetűek.
- (c) Mindebből pedig *nem következik az*, hogy ha létezik elme, akkor annak okoznia kell valamit a fizikai világban, mert lehetséges olyan álláspont (akkor is, ha nem szeretjük), amely szerint mind a mentális, mind a fizikai szféra okságilag zárt, ugyanakkor a mentális állapotok nem okoznak fizikai állapotokat, és a fizikai állapotok nem okoznak mentális állapotokat, azaz a mentális és fizikai események vagy eseménysorozatok viszonya két, egymással tökéletes szinkronba állított óra működéséhez hasonlítható. Ez pedig nem más, mint a pszichofizikai parallelizmus álláspontja, amelyet például Leibniz képviselt.

Mindazonáltal leginkább nem a következtések helytelenségét kifogásoljuk, hanem azt, hogy Ullmann azt a látszatot kelti, mintha a mentális okozásról való beszéd valamiféle tárgyontológia reflektálatlan elfogadásának a következménye lenne. Való igaz: a kortárs analitikus filozófusok döntő többsége úgy gondolja, hogy az elménk oksági viszonyban áll fizikai környezetünkkel. Bizonyos mentális állapotokat fizikai események okoznak, bizonyos fizikai eseményeket pedig mentális állapotok idéznek elő. Csakhogy – és ez a fő kifogásunk – e meggyőződésünknek *semmi köze nincsen* semmiféle tárgyontológiához, mivel e meggyőződésünk az egyik legtermészetesebb, legegyszerűbb emberi meggyőződés. A lehető legtermészetesebb dolog azt gondolni, hogy amikor valaki kalapáccsal az ujjára üt, akkor az illető fájdalmát (mentális állapotát) a kalapács ütése (fizikai esemény) okozza. Ugyancsak a lehető legtermészetesebb dolog azt gondolni, hogy ha valaki félelmében elfut a fogorvosa elől, akkor az illető futását (fizikai esemény) az illető félelme (mentális állapota) okozta.

2. Ullmann tanulmányának *Jelentés, referencia, leképezés* című részében – többek között – leszögezi:

Az analitikus filozófiában – természetes összhangban a hallgatólagosan uralkodó fizikalista tárgyontológiával – a jelentés referenciaelmélete vált uralkodóvá. [...] Ám hogy ez [mármint az, hogy a jelentést a referencia felől kell megközelíteni] lehetséges legyen, el kell fogadnunk azt a feltevést, hogy a valóság olyan szilárd és eleve adott szerkezettel rendelkezik, hogy a nyelv szerepe alapvetően redukálható a referencialításra. A realizmus-antirealizmus szembeállításából itt most a következő érdekes: a realizmus alaptétele nem pusztán az, hogy van külvilág, hanem az, hogy a világnak, a dolgoknak stb. van egy adott szerkezete (ezt a természettudomány majd feltárja). Következésképp a nyelvet akkor használjuk helyesen, ha az igazság meghatározott feltételei mellett jól referálunk erre a dolghalmazra, és nem kerülünk ellentmondásba a természettudomány meggyőződéseivel (Ullmann 2006. 233–234).

Bevalljuk, hogy nem egészen világos számunkra, mit ért Ullmann a „jelentés referenciaelméletén”. Ha a Kripke–Putnam-féle direkt (okági) referenciaelméletre gondol, akkor sajnos azt kell mondanunk, félreérti ezen elmélet motivációját, ugyanis sem Kripke (1972/2007), sem Putnam (1981) nem képviselnek „fizikalista tárgyontológiát”. Ők egészen más úton, a leírások és a tulajdonnevek szemantikai tulajdonságainak vizsgálata felől jutottak el ehhez az elmülethez, és bár a referencia kauzális elméletét valóban felhasználják fizikalisták is, elfogadása vagy elutasítása *független* a fizikalista ontológiától. *Nota bene*: Kripke a direkt referencia elméletében szereplő „merev jelölő” fogalmát a test és az elme dualizmusának igazolására próbálta meg felhasználni. Putnam modelleméleti érve pedig – amelyben a kauzális elméletet hívja segítségül azon tézis alátámasztásához, hogy „a referenciát nem az határozza meg, ami a fejünkben van” – éppen a metafizikai realizmus ellen irányul, amelyet (részben) azzal határoz meg, hogy „a világot az elmefüggetlen tárgyak rögzített totalitása alkotja”, illetve, hogy „az igazság valamilyen típusú korrespondencia (leképezés) a szavak és a külvilág dolgai között” (Putnam 1981. 49).

De akkor is kénytelenek vagyunk ellentmondani Ullmann-nak, ha úgy gondolja, hogy a referencia szemantikai szerepének hangsúlyozása általában vezet fizikalista elköteleződéshez. Mint azt Higginbotham (2006) megállapítja: az igazság és a referencia annyira alapvetőek, hogy a természetes nyelvi jelentés-fogalom bármilyen teoretikus elemzése során *kiindulópontul* kell szolgálniuk – a kettő közötti kapcsolatot a mondatjelentés igazságfeltételes (vagy antirealista nézőpontból: állíthatósági feltételekkel jellemezhető) értelmezése biztosítja.

Lássuk be, mindebből megint csak nem következik az, hogy egy bizonyos metafizikát kellene elfogadnunk. Az igazság (vagy igazolt állíthatóság) és a re-

ferencia központi szerepét a jelentésemleletekben az antirealisták sem tagadják (Dummett 1991/2000. 74–94). Már a realizmus–antirealizmus-vita kapcsán is világosan kiderül: a fizikalista vagy bármilyen más ontológia, illetve a „szilárd és eleve adott szerkezetű” valóság koncepciójának elismerése függetlenek a referenciális (igazságfeltételes) szemantika elfogadásától. Talán a legjobb példa erre maga az igazságfeltételes szemantika kidolgozója, Frege (1918/2000), akit egészen abszurd volna fizikalistának tekinteni, hiszen platonista volt, és három létartományt ismert el (mentális képzetek – fizikai világ – absztrakt entitások). Sőt, a jelentés ilyen megközelítése minden további nélkül összeegyeztethető például azzal a nézettel, hogy a fizikai tárgyak az elme által érzetadatokból konstruált létezők (antirealista fenomenalizmus); vagy azzal az elmélettel, hogy léteznek redukálhatatlan társadalmi-institucionális tények, amelyekre referálni tudunk. Mi több, a kortárs szemantikában a jelentések formális reprezentációjára alkalmazott modelleméleti keret elfogadása is nyitva hagyja azt a kérdést, hogy mik legyenek a modell elemei (lásd például Kálmán–Rádai 2001. 17). Világos, hogy ezen keretek között beletartozhatnak a modellbe lelki jelenségek, társadalmi entitások, folyamatok, események is, így akár egy Whitehead-féle folyamatontológiát is modellezhetnénk úgy, hogy közben elfogadjuk a referencialitás alapvető szerepét a nyelvi jelentés elemzésében.

3. Ullmann a tanulmány *Intencionalitás* című részében a következőképpen fogalmaz:

Az intencionalitás analitikus filozófiában meghonosodott fogalma [...] a következőt jelenti: vannak az elmében olyan hitek, meggyőződések, mentális állapotok, amelyek nem képezhetők le fizikai összefüggésekre, ugyanakkor valamilyen módon a világra vonatkoznak. A pszichikai fenomének nem redukálhatók a fizikai fenoménekre, különbségük – immár Brentano szerint – abban áll, hogy a pszichikai fenomének mindig irányulnak valamire, vagyis intencionálisak, a fizikai fenomének ellenben nem irányulnak semmire. Ez úgy jelenik meg a kortárs analitikus elmefilozófiában, hogy a hitekről szóló állítások nem képezhetők le nem-intencionális, vagyis fizikai leírásra. Egyszerűbben: az elme nem redukálható az agyra (Ullmann 2006. 235).

Ullmann fenti állításai nagyon félrevezetőek, de mert fontos dolgot érintenek, érdemes tisztázni a problémát.

Mindenekelőtt szögezzük le: az analitikus filozófusokat az intencionalitás jelensége *több okból* is foglalkoztatja. Íme a legfontosabb, intencionalitással kapcsolatos problémák:

- (a) Hogyan fest az intencionális állapotok szerkezete?
- (b) Lehetséges-e az intencionalitás jelenségére naturalista magyarázatot adni, s ha igen, akkor hogyan?
- (c) Milyen viszonyban áll az intencionalitás a nyelvvel? Melyik az elsődleges?
- (d) Milyen viszonyban állnak az intencionális állapotaink a környező világgal? Függenek-e intencionális állapotaink tartalmi attól, hogy milyen világban élünk? Precízebben: lehetséges-e egymástól releváns módon különböző lehetséges világokban ugyanazzal az intencionális állapottal rendelkezni, vagy sem? (Ez az internalizmus és externalizmus vitája.)
- (e) Milyen viszonyban áll az intencionalitás a fenomenális tudatossággal? Vannak-e olyan mentális állapotok, amelyek nem intencionálisak, de fenomenális-tudatosak? Illetve: vannak-e olyan mentális állapotok, melyek nem fenomenális-tudatosak, de intencionálisak?

Visszatérve az idézett passzusra: Ullmann csúsztat is, téved is. Egyrészt csúsztat, mert úgy állítja be, mintha az analitikus filozófusokat *kizárólag* annyiban érdekelné az intencionalitás, hogy az redukálható-e agyállapotokra, és hogy lehetséges-e valamiféle naturalista magyarázatot adni rá. Ez azonban nincs így, ugyanis – mint azt fentebb leírtuk – az analitikus filozófusok több aspektusból is vizsgálják az intencionalitás jelenségét, és nemcsak a (b) kérdéssel foglalkoznak.

Másrészt pedig téves az állítása, mert a legtöbb kortárs analitikus filozófus – Roderick Chisholm és persze Brentano álláspontjával szemben – úgy gondolja, hogy az intencionalitás (mint az elme azon tulajdonsága, hogy valamire irányul) igenis magyarázható naturalisztikusan. Erre nézvést jól kidolgozott elméletek állnak rendelkezésre. Gondoljunk Jerry Fodor oksági elméletére (1987), Ruth Garrett Millikan teleologikus elméletére (1984), Dennett fikcionalista elméletére (1971/1998), és még folytathatnánk.

Néhány sorral később még zavarosabbá válik a tanulmány gondolatmenete:

Két álláspont vezet itt [mármint az intencionalitás kapcsán] a vitákat: a reduktív fizikalizmus (az elmeállapotok, tehát az intencionalitás is, végső soron visszavezethető a fizikai összefüggésekre); és a nem-reduktív fizikalizmus, amely nem tagadja ugyan a fizikai tudományok végső magyarázóértékét, de a mentális állapotokat nem tartja teljesen visszavezethetőnek ezekre. Az elsővel most nem foglalkozom, a második már érdekesebb. Ugyanis a második álláspont jól tükrözi az analitikus filozófia zavarát, ahogy közeledik a tudat és a tudatosság problémájához. Minden jel arra mutat, hogy az ún. elme, de legalábbis a mentális állapotok, nem vezethetőek vissza maradéktalanul fizikai történésekre (Ullmann 2006. 235).

Hogy mi tükrözi az analitikus filozófia zavarát, és mi nem, az valóban lehet kérdés – de hogy nem az, amit Ullmann mond, az egészen bizonyos. Állítása sokkal inkább minket hoz zavarba, mivel e passzusban Ullmann ellentmond korábban idézett szavainak. Míg korábban azt állította, hogy a kortárs analitikus elmefilozófiában az intencionalitás nem redukálható fizikai entitásokra (az agyra), addig e passzusban azt állítja, hogy a kortárs elmefilozófiában *éppen arról* folyik a vita, hogy redukálhatóak-e vagy sem a mentális állapotok (így az intencionálisak is) fizikai entitásokra.

De spongyát rá! Vannak komolyabb kifogásaink is. Egyrészt nem tartjuk különösebben szerencsésnek úgy meghatározni a reduktív fizikalizmus fogalmát, hogy az „a mentálist *visszavezeti* a fizikaira”, ugyanis a „redukción” szó épp azt jelenti, hogy „visszavezetés”, következésképpen e meghatározás nem túl informatív. Helyesen: a reduktív fizikalizmus azt állítja, hogy a mentális állapotok típusai (a mentális tulajdonságok) *azonosak* neurofiziológiai állapottípusokkal (neurofiziológiai tulajdonságokkal). A redukción ugyanis mindig *azonosítás*: amennyiben *a* entitást ontológiai értelemben redukálom *b* entitásra, úgy *a*-t azonosítom *b*-vel. Ennek értelmében nem-reduktív fizikalizmuson azt a nézetet értjük, amely szerint a mentális tulajdonságok (mentális állapotaink típusai) *nem azonosak* fizikai (neurofiziológiai) tulajdonságokkal, vagyis a „mentális” és a „fizikai” mint tulajdonságok különbözőek (ettől nem-reduktív), noha a „fizikai” határozza meg a „mentális”-t (ettől fizikalizmus).

Ullmann tehát pontatlanul fogalmaz, de a fenti idézet utolsó mondata szíven ütött minket. Ha ugyanis tárgyilagosak akarunk maradni, akkor – szemben azzal, amit Ullmann mond – igenis el kell ismernünk, hogy ha nem is minden, de számos (empirikus) jel épp arra utal, hogy a (reduktív vagy nem-reduktív) fizikalizmus a helyes álláspont. Beyersteint (1991) követve több empirikus evidenciát hozhatunk fel az agy és az elme közt fennálló aszimmetrikus függésre. Először is, a filogenetikus bizonyítékok azt támasztják alá, hogy a törzsfajlás során az agy (idegrendszer) komplexitásának foka és az egyes fajok kognitív képességeinek szintje egyenes arányban állnak egymással. Corliss Lamont szerint „minél nagyobb az agy (agykéreg) mérete az élőlény testéhez viszonyítva, és minél komplexebb annak struktúrája, annál magasabb rendű és sokoldalúbb az élőlények életformája” (Lamont 1990. 63). Másodsor, a mentális képességek az agy fejlődésével jelennek meg: az agy fejlődését akadályozó tényezők akadályozzák a mentális fejlődést is. Harmadsor, a klinikai bizonyítékok szerint az egyes balesetek, mérgezések, betegségek vagy alultápláltság okozta agysérülések a mentális funkciók visszafordíthatatlan károsodásához vezetnek. Negyedszer, a legerősebb empirikus bizonyítékot az agy közvetlen elektromos vagy kémiai stimulációja nyújtja, amelynek során érzetek, emlékek, vágyak és egyéb mentális állapotok jönnek létre a páciensekben.

Ha kacskaringósan is, de Ullmann gondolatmenete végül oda konkludál, hogy az analitikus filozófiában „elsősorban az nincs tisztázva, hogy vajon milyen



perspektívából nézzük az intencionalitást, holott itt az első személyű perspektíva és a harmadik személyű perspektíva különbsége nagyon jelentőssé válik” (Ullmann 2006. 236). Ullmann-nak tökéletesen igaza van abban, hogy az első és harmadik személyű perspektíva közötti különbség alapvető fontosságú. Abban azonban nincsen igaza, hogy ez tisztázatlan volna az analitikus filozófiában.

Ullmann konklúziója azért hamis, mert figyelmen kívül hagyja azt a tényt, hogy az intencionalitással kapcsolatban az analitikus filozófiában nem egy, hanem *több* probléma is felmerül – ahogy azt fentebb pontokba szedve is bemutattuk. Ezek közül egyes problémák elemzése – mint például az intencionalitás naturalizálásának problémája – elsősorban egyes szám harmadik személyű perspektívát igénylő munka, míg az intencionális állapotok intencionális szerkezetének feltárása elsősorban egyes szám első személyű perspektívából történik (lásd például Crane 2001. 1. fejezet).

4. Ullmann a tanulmány *Tudat* című részében ekképpen ír:

Tömören összefoglalva a helyzetet: mivel az analitikus filozófia kezdeteinél a kiindulópont nem más volt, mint az a meggyőződés, hogy a fizikai természet tudományos leírása adja meg a valóság megismerésének végső kulcsát, valamint hogy nincs más, csak ami ennek tárgya lehet, csak az objektív, fizikai létezés van, ezért egyáltalán nem meglepő, hogy amikor tekervényes utakon az analitikus filozófia számára is elkerülhetetlenné válik a szembenézés a tudat problémáival (a kognitív tudományok fejlődésének kényszere alatt), akkor csak az elutasítás marad (Ullmann 2006. 236–237).

A fentiekhez két hozzáfűznievalónk van. Az első: Ullmann állításával szemben a kortárs analitikus filozófia ipari méretekben foglalkozik a fenomenális tudat problémájával. A vita nagyjából a következőképpen fest: a legtöbb kortárs analitikus filozófus egyetért abban, hogy a fizikalista/materialista ontológia számára a fenomenális tudat jelenti a legfőbb kihívást. Innen ágaznak el a különböző vélemények. Míg az egyik tábor úgy véli, hogy a fenomenális tudatosságra alapozva konkluzív módon érvelhetünk a fizikalizmus ellen, addig a másik úgy, hogy az antifizikalista érvek nem konkluzívak, és igenis magyarázható a fenomenális tudatosság a fizikalizmus ontológiáján belül. E két tábor vitája az analitikus filozófia talán legnagyobb élő vitája, következésképpen téves azt állítani, hogy az analitikus filozófia elutasítja a fenomenális tudatossággal való szembenézést.

Persze mindezzel nyilván Ullmann is tisztában van. Máskülönbem nem ismerette volna (tegyük hozzá: teljesen korrekt módon) David Chalmers álláspontját (2006. 237–239), amely szerint a fizikalizmus nem képes megoldani a fenomenális tudat problémáját. Mi csak azt nem értjük, hogy Ullmann miért állítja mégis, hogy az analitikus filozófia elutasító gesztussal viseltetik a fenomenális tudatos-

ság iránt? Talán azt gondolja: *egyedül* Chalmers foglalkozik a fenomenális tudatossággal? Vagy azt, hogy *egyedül* Chalmers antifizikalista? Akárhogy is: Ullmann egy, a fenomenális tudatossággal foglalkozó, antifizikalista, analitikus filozófus munkáján keresztül véli megmutatni azt, hogy az analitikus filozófia elutasító gesztussal viseltetik a fenomenális tudatosság problémájával szemben. Meglehetősen furcsa eljárás.

A második hozzáfűznivalónk a következő: ha Ullmann azt akarja állítani, hogy az analitikus filozófiában a fenomenális tudatosság azért okoz problémát, mert az analitikus filozófia alapvetően fizikalista beállítottságú, akkor igaza van. Azt azonban már helytelenül teszi, hogy egyetértően idézi, és „mély értelmű megfogalmazás[nak]” tekinti Szigeti Attila ama passzusát, amely szerint „*nem lehet a tudatosságot egy olyan módszerrel vizsgálni, amelynek alapító gesztusa pontosan a tudatosság kiiktatása volt*” (Szigeti 2002. 103, idézi Ullmann 2006. 238, kiemelés az eredetiben).

Sejtelmünk sincs, hogy Szigeti (és Ullmann) miből gondolja, hogy létezik olyan módszer, amelynek „alapító gesztusa” a tudatosság kiiktatása lenne. Ugyanakkor az valóban fontos kérdés, hogy az analitikus filozófia miért fizikalista beállítottságú. Ennek több oka van (lásd cikkünk előző részét), témánk szempontjából azonban a következő a fontos: az analitikus filozófusok többsége azért fizikalista, mert úgy gondolja: kizárólag akkor lehetséges plauzibilisen magyarázni a mentális okozást, ha a mentális állapotokat/eseményeket fizikai állapotokkal/eseményekkel azonosítjuk. Íme a fizikalizmus mellett felhozott érv általános sémája (Lewis 1966; Armstrong 1968):

- (a) A mentális állapotok okozhatnak fizikai eseményeket. (Azaz létezik mentális okozás.)
- (b) A fizikai világ okságilag zárt: minden fizikai eseménynek van elégséges fizikai oka. (Azaz: minden  $e$  fizikai eseménynek van  $e'$  fizikai oka, úgy, hogy  $e'$  önmagában elégséges ahhoz, hogy  $e$ -t létrehozza, összhangban a megfelelő fizikai törvényekkel.)
- (c) A cselekedetek nem túldetermináltak fizikai és mentális okuk által. (Azaz: az emberi cselekvések esetében nem áll fenn oksági túldetermináció, vagyis egy cselekedetet nem egy fizikai és egy attól függetlenül létező mentális állapot okoz olyan módon, hogy mind a fizikai, mind pedig a mentális állapot önmagában elégséges a kérdéses cselekedet létrehozásához.)

Következésképpen:

- (d) A mentális okok = fizikai okok.

Következésképpen:

- (e) A mentális állapotok = fizikai állapotok.

Tehát a fizikalizmus szerint, ha egy mentális állapotnak van fizikai hatása, akkor az a mentális állapot fizikai állapot. Nem két oka van egy cselekedetnek, egy fizikai és egy mentális (amelyek együttesen túldeterminálják a cselekedetet), hanem *egyetlenegy*. Egy cselekedet mentális oka ugyanis nem más, mint a kérdéses cselekedet fizikai oka.

Tény, hogy a kortárs analitikus elmefilozófusok jelentős része fizikalista – de ez nem egy dogmatikus előfeltevés reflektálatlan elfogadásának eredménye, hanem egy *érv* győzte meg az analitikus filozófusok zömét. Nevezetesen: ha hiszünk abban, hogy mentális állapotainknak lehetnek fizikai hatásai, továbbá ha hiszünk abban is, hogy minden fizikai eseménynek van elégséges fizikai oka, és ha elutasítjuk azt, hogy a cselekedeteink túldetermináltak mentális és fizikai okaik által, akkor el kell kötelezni magunkat amellelt, hogy a mentális állapotok fizikai állapotok.

De hogyan kerül a képbe a fenomenális tudat? Válasz: a fenomenális tudat problémája a mai formájában azután merült fel, hogy az elmefilozófiában uralkodóvá vált a fizikalizmus, vagyis az a nézet, hogy a mentális állapotok fizikai állapotok. Thomas Nagel (1974/1997) és utána sokan mások is amellelt érveltek, hogy a fizikalizmus ontológiáján belül nehézségekbe ütközik a fenomenális tudatosság magyarázata. A probléma: nehéz (vagy lehetetlen?) megérteni azt, hogyan lehetnek azonosak fizikai állapotokkal a fenomenális karakterrel rendelkező mentális állapotaink. Hogyan lehet egy lényegét tekintve szubjektív természetű entitás az objektív világ eleme?

Összefoglalva: Ullmann-nak igaza van abban, hogy a fenomenális tudatosság jelensége speciálisan a fizikalizmus számára jelent kihívást. Ha valaki történetesen szubsztanciadualista, annak a fenomenális tudatosság magyarázata nem jelent problémát, mivel nem tekinti a fenomenális mentális állapotokat fizikaiaknak. Csakhogy Ullmann nem veszi figyelembe, hogy a fizikalizmus elfogadása nem dogmatizmus, hanem sokak szerint az egyetlen lehetőség, amennyiben plauzibilisen magyarázni akarjuk a mentális okozást.

5. Ullmann tanulmányának *Józan ész* című részében azon meggyőződésének ad hangot, hogy az analitikus filozófia alapvető módszertani előfeltevéséi közé tartozik a józan észbe vetett feltétlen, de legalábbis nemigen megkérdőjelezett bizalom. Ullmann ezt a következőképpen fogalmazza meg:

A hallgatólagos előfeltevésekkel szembeni érzéketlenség leginkább abban a formában jelenik meg az analitikus filozófián belül, hogy közvetlenül vagy közvetve a józan észre hivatkoznak. Ezt és ezt a meggyőződést nem kell felülvizsgálni, sőt még csak megvizsgálni sem, mert a „józan ész” igazolja. [...] Alighanem az angolszász hagyománynak köszönhető az analitikus filozófia rendkívül erős bizalma a józan ész mindenféle formáiban.

Ez nem csupán az *ordinary language philosophy* irányzatában figyelhető meg, hanem – talán a tiszta logikai kutatások kivételével – szinte az analitikus filozófia minden változatában (Ullmann 2006. 239–240).

Ullmann téved. Azt nem vitatjuk, hogy az analitikus filozófiának volt olyan korszaka – jelesül a természetes nyelv filozófiája –, amelyben kiemelt szerepet játszott a józan ész. Azt sem vitatjuk, hogy az analitikus filozófiának mindig voltak, és ma is vannak olyan képviselői, akik kitüntetettnek tekintették/tekintik a józan észet. De azt vitatjuk, hogy az analitikus filozófiát általánosan jellemezné a józan észbe vetett „rendkívül erős” bizalom.

Nézzük először az analitikus metafizika és a józan ész viszonyát. Peter Frederick Strawson (1959) nyomán különbséget teszünk a metafizika két típusa, *deskriptív* és *revizionista* metafizika között. A deskriptív metafizikának az a célja, hogy feltárja és leírja a világról alkotott fogalmi sémánkat, amellyel mi, emberek, valamennyien rendelkezünk. Ez az elménkben levő fogalmi struktúra vagy séma nem változik az időben, teljesen független attól, hogy milyen nyelven beszélünk, illetve gondolkodunk, és annyira mélyen gyökerezik bennünk, hogy nem lehetséges tőle elvonatkoztatnunk, bárhogy is gondolkozunk, és bármilyen tudományos elméletet fogadunk is el egyébként. E bennünk rejlő fogalmi struktúra pedig elsősorban a józan ész meggyőződéseiben tükröződik; Strawson szerint például abban, hogy valamennyien úgy gondoljuk, hogy alapvetően fizikai tárgyak és emberi személyek léteznek. A deskriptív metafizikát tehát valóban jellemzi a józan észbe vetett bizalom.

Csakhogy a kortárs metafizika – legalábbis túlnyomórészt – revizionista, és elutasítja a deskriptív metafizikát. Ennek az elutasításnak az egyik legfőbb oka az a felismerés, hogy a természetes intuitív meggyőződéseink (vagy: józan észben alapuló hiteink) sok esetben inkonzisztensek egymással, és így egyszerre mindegyikük nem lehet igaz. Egyetlen példa: az egyik legalapvetőbb meggyőződésünk szerint a fizikai tárgyak (emberek, növények, autók, könyvek, házak stb.) *változnak az időben*. Másik alapvető meggyőződésünk szerint a fizikai tárgyak *változásaik ellenére ugyanazok maradnak*. E két meggyőződésünk között feszültség van: ugyanis ha egy fizikai tárgy megváltozik, akkor más tulajdonságokkal rendelkezik, mint korábban, és a korábbi időpontban *F* tulajdonsággal rendelkező tárgy – a Leibniz-törvény értelmében – nem lehet azonos a későbbi időpontban *nem-F* tulajdonsággal rendelkező tárggyal. Ennélfogva – mivel egy konzisztens elmélet kidolgozása a célunk –, vagy az egyik, vagy a másik természetes meggyőződésünket fel kell adnunk. És valóban: a kortárs (a fizikai tárgyak változásával kapcsolatos) metafizikai viták – némileg leegyszerűsítve – éppen akörül forognak, hogy e két alapvető (józan észben alapuló) meggyőződésünk közül melyiket kell feladni. (A deskriptív és revizionista metafizikákról, valamint a változás problémájáról lásd bővebben Tózsér 2009.) E példa világosan mutatja, hogy nem igaz Ullmann sommás

megállapítása, miszerint az analitikus filozófiában a józan ész meggyőződéseit „nem kell felülvizsgálni, sőt még csak megvizsgálni sem”.

De vonatkoztassunk el a metafizikától. Vegyük például Quine-t. A quine-i holisztikus igazolás- és jelentéskoncepciót elfogadó analitikusok szerint egyetlen állítás sem mentes a revízió lehetőségétől. Következésképpen ők sem viseltetnek erős bizalommal a józan ész iránt. Keith Campbell, aki az analitikus filozófia és a józan ész viszonyát három történeti fázisra osztja, ekképp fogalmaz: „[a] Quine-féle fázisban [...] a józan ész legjobb esetben is csak szerény szerephez jut: legfeljebb a teljes elméleti sémák számára biztosít néhány összetevőt, önmagában véve nincs presztízse. Amiatt pedig egyetlen séma sem érdemel kiemelt figyelmet, hogy a józan ész meggyőződéseit tartoznak bele” (Campbell 1988. 165).

Ullmann tévedését azzal szemléltethetjük a legjobban, ha bemutatunk néhány olyan álláspontot, amelyeket analitikus filozófusok képviseltek, míg a józan ész számára egészen bizonyosan elfogadhatatlanok. Nagyon sok példát említhetnénk, de álljon itt csak négy:

- (a) Egyes analitikus filozófusok szerint (Feyerabend 1963; Rorty 1965; 1970; 1979. 2. fejezet; Paul Churchland 1981/1991; Patricia Churchland 1986; Stich 1983; Ramsey 1991; Rey 1983; 1988 stb.) – *szemben* a józan ész álláspontjával – nem léteznek mentális állapotok, vagyis hitek, vágyak, félelmek stb.
- (b) David Lewis – *szemben* a józan ész álláspontjával – azt állítja: „megszámálhatatlanul végtelen sok számár és proton és tócsa és csillag és a Földhöz nagyon hasonló bolygó és Melbourne-höz nagyon hasonló város és Önhöz nagyon hasonló személy” létezik (Lewis 1986. 133).
- (c) David Benatar – figyeljünk a címre! – *Better Never to Have Been* (2006) című könyvében amellet érvel, hogy – *szemben* a józan ész álláspontjával – megszületni és tudatos lényként létezni kivétel nélkül mindig rossz dolog, és ezért az volna a legjobb, ha az emberek nem vállalnának gyerekeket, és az emberiség fokozatosan kipusztulna, és egyáltalán, a Földön megszűnne minden tudatos életforma.
- (d) Nick Bostrom (2003) – ugyancsak *szemben* a józan ésszel – amellet érvel, hogy rendkívül valószínű, hogy a világ, amelyben élünk, valójában a Mátrixhoz hasonló számítógépes szimuláció.

Összefoglalva: nem igaz, hogy az analitikus filozófiában a józan észnek episztemikus privilégiumai volnának. Az analitikus filozófia radikálisan különböző nézeteket és módszereket tartalmazó, nagyon sokszínű szellemi alakzat, amelyről nagyon félrevezető teljesen általánosan kijelenteni, hogy a józan észnek kiemelt szerepet szán.

## II. A KONTINENTÁLIS FILOZÓFIA VÉDELME EGY FENOMENOLÓGUS NÉZŐPONTJÁBÓL

A tanulmány második felében Ullmann Tamás arra vállalkozik, hogy – nagyjából az azt megelőző rész pontjait (fordított sorrendben) követve – elmagyarázza, hogy miképp vélekednek a fenomenológusok ugyanezekről a dolgokról, és hogy mennyivel plauzibilisebb elképzeléseket vallanak, mint analitikus kollégáik. Véleményünk szerint Ullmann ígyekezetét nem koronázza siker.

1. Ullmann *A józan ész és a metafizika meggyőződéseivel való leszámolás* című részben a fenomenológusok által követett, vagy követendő eljárást a következőképpen írja le:

A fenomenológia „alapító gesztusa” nem pusztán módszertani bibelődés, hanem annak felismerése, hogy a gondolkodás legelevenebb szükséglete a hallgatolagos előítéletekkel való leszámolás. Itt nem csupán a metafizikai hagyomány előítéleteiről van szó, hanem a kulturálisan leülepedett, a mindennapi meggyőződéseinket is meghatározó előítéletekről. Vagyis a józan ésszel és a metafizikával egyaránt szakítani kell. Ez radikális szemléletváltást jelent. Nem lehet mellette érvelni, mert minden érvelés valamilyen meggyőződésre támaszkodik: a meggyőzések feladása mellett nem szólnak érvek, hiszen minden lehetséges érv valamilyen meggyőződésre épít. Az előítéletek feladásával a fenomenológia (és a kontinentális filozófia más ágai) a tapasztalati horizont előzetes „megtisztítására” szólítanak fel (Ullmann 2006. 241).

A legtöbb fenomenológus valószínűleg elfogadja Ullmann állításait, mi azonban elhamarkodottnak tartjuk az efféle eljárást.

Különösen árulkodónak tartjuk, hogy Ullmann a józan ésszel való „leszámolásról” beszél. Ezzel nyilvánvalóan azt előfeltételezi, hogy a józan ész, úgy ahogy van, félrevezető és hamis képet fest a világról és benne önmagunkról. De mi ez, ha nem dogmatizmus? Nem dogmatizmus-e vajon minden előzetes vizsgálat *nélkül* kijelenteni, hogy a józan észben alapuló meggyőződéseink „a tapasztalatot torzító” meggyőzések? (Ullmann 2006. 242)

Mi a „leszámolásnál” lényegesen szerényebb és egyúttal célravezetőbb módszert ajánlunk. Íme: minden egyes filozófiai probléma kapcsán (a) tegyük explicitté természetes vagy józan észben alapuló meggyőződéseinket, és (b) vizsgáljuk meg, hogy azok tarthatóak-e vagy sem. Ahogy az előző pontban írtuk: természetes meggyőződéseink gyakran inkonzisztensek, vagyis együttesen nem állhatnak fenn. Vizsgáljuk meg tehát külön-külön, hogy e meggyőzések közül melyek azok, amelyekhez valamilyen elméleti megfontolásból ragaszkodnunk kell, és melyek azok, amelyekhez nem, vagy csak kevésbé.

Úgy gondoljuk, hogy e munka során – anélkül, hogy fenomenológusok volnánk – igenis nagy „gyakorlatot [lehet] szerezni a gondolati »rögzültségek« feloldásában” (Ullmann 2006. 242).

2. A tanulmány *A tudat* című része kapcsán először is azt jegyeznénk meg, hogy Ullmann sajnos nem tesz különbséget a tudat mint fenomenális tudat, és a tudat mint valamiről való tudat fogalma között. A tudat e két fogalma folyamatosan keveredik a szövegben, igencsak megnehezítve a megértést. Íme:

Az Én problémájánál is mélyebbre nyúlunk az, ami az analitikus elmefilozófiában úgy jelenik meg, mint szubjektív, fenomenális tudás: ez tulajdonképpen annyi, hogy átélem a tapasztalatot és egyszersmind tudok is róla. A tudat problémája nem az Én problémája, hanem az önmagáról tudó tudaté (Ullmann 2006. 242).

Eltekintve a permanens terminológiai zavartól, Ullmann e részben fontos kijelentést tesz, amely szerint a fenomenológia perspektívájából nézve a test–elme-probléma nem alapvető filozófiai probléma. Így ír:

[A] fenomenológia szerint a tudat nem lélek a testtel szemben, illetve [...] nem elme az aggyal szemben. Ez a szembeállítás ugyanis már feltételez egy bonyolult szemantikai mezőt és egy hallgatólagos metafizikát. A test és az elme külön-külön és egymásra vonatkoztatva is csak a tudat felől érthető meg, vagyis a tudat megelőzi ezt a megkülönböztetést: a tudat az a pont, ahonnet láthatóvá válnak az ilyen [...] „konstituált” értelemalakzatok (Ullmann 2006. 243).

Fenomenológus kollégák szájából gyakran halljuk, hogy a tudat mi mindent *előz meg*, ezért most sem lepődünk meg különösebben azon, hogy Ullmann szerint a test és az elme megkülönböztetését is megelőzi. Csakhogy az a tézis, hogy „a tudat *megelőzi* ezt a [test és az elme] megkülönböztetés[é]t”, kétféleképpen érthető, és Ullmann nem különbözteti meg e tézis két különböző értelmét.

Az egyik értelemben azt jelenti Ullmann tézise, hogy a test–elme-probléma felvetéséhez már rendelkezniünk kell a test és az elme – többé-kevésbé világos – fogalmával, vagyis e probléma megértése előfeltételezi azt, hogy tudatában legyünk a test és elme – többé-kevésbé világos – fogalmának. Ezt nevezhetnénk a tudat *episztemológiai* értelemben vett elsőbbségének. (Egy csimpánz számára nyilván nem merül fel a test–elme-probléma, a csimpánz ugyanis nincs tudatában a megfelelő fogalmaknak.) Egyszóval: abban az értelemben *előzi meg* a tudat a test és az elme fogalmának megkülönböztetését, amennyiben bármely más fogalmi megkülönböztetést is megelőz.

A másik értelemben viszont azt jelenti Ullmann tézise, hogy a tudat *ontológiai* elsőbbséggel rendelkezik a testhez és az elméhez képest. És Ullmann valószínűleg erre is gondol, hiszen úgy fogalmaz: a test és az elme „*konstituált* értelemalakzatok”. E szerint: a test és az elme annyiban és csak annyiban léteznek, amennyiben valamilyen tudat által konstituált entitások. Ez azonban – mondjunk sem kell – vagy a Berkeley-féle idealizmus, vagy a transzcendentális idealizmus elfogadását implicálja. Ez nem feltétlenül baj, de azért nem ártott volna, ha ezt Ullmann explicitté teszi, ugyanis az idealizmusnak mindkét változatával szemben komoly ellenvetések hozhatók fel.

Összefoglalva: Ullmann tézise vagy triviálisan igaz, vagy nem triviális ugyan, de vitatható.

3. Ullmann tanulmányának *Intencionalitás* című része egy fontos tézissel kezdődik:

Nincs tudat intencionalitás nélkül és – ami az analitikus irányzattal szemben lényeges – nincs intencionalitás tudat, azaz fenomenális tudatosság nélkül (Ullmann 2006. 244).

Az előző pontban már utaltunk rá, hogy Ullmann nem különbözteti meg a tudat mint fenomenális tudat, illetve a tudat mint valamiről való tudat fogalmakat, de ezen most felülemelkedünk, ugyanis nagyon valószínű, hogy a következőt akarja állítani: a fenomenológia szerint (a) minden intencionális mentális állapot egyúttal fenomenális-tudatos, és (b) minden fenomenális-tudatos mentális állapot egyúttal intencionális. Más szavakkal: nincsenek olyan mentális állapotok, melyek fenomenális-tudatosak ugyan, de nem intencionálisak, és olyan mentális állapotok sincsenek, melyek intencionálisak ugyan, de nem fenomenális-tudatosak. (Ullmannhoz hasonlóan érvel még például McGinn 1991; Chalmers 2006; Gallagher–Zahavi 2008. 115 skk.) Ullmann tézise tehát egy valóban lehetséges (sőt, igen markáns) álláspont, amely éppenséggel igaz is lehet (e sorok írói közül T. J. egyetért vele), de azt is látni kell: mind az (a)-val, mind a (b)-vel szemben komoly ellenvetéseket lehet felhozni.

Ugyanakkor nagyon elégedetlenek vagyunk azzal, ahogyan Ullmann a későbbiekben az intencionalitás fogalmát jellemzi:

Az egyes szám első személy perspektívájából feltárulkozó tudat nem elszigetelt edény, amelynek alkalmoszerű viszonya „támad” más dolgokkal, amelyek valahol „kívül” vannak: a tudat intencionális tudat és ezáltal nyitottság. Az intencionális tárgy nem úgy áll szemben a tudattal, mint egy objektum a szubjektummal. Nem két egymástól eleve elválasztott dolog lép itt egymással kapcsolatba, hanem az intencionális tudat eleve tárgyra irányul [...], a tárgy pedig nem választható le a rá irányuló tudatról (ami



persze nem jelenti azt, hogy feloldódna valamiféle idealista látszatlétben). [...] A fenomenológia szerint az intencionalitás pontosan az a jellemzője a tudatnak, hogy nem kilép a tárgyhoz, hanem eleve minden formájában már a tárgynál van. Éppen az intencionalitás fogalmával sikerül meghaladni a „tudat-interioritás” karteziánus metafizikáját: hiszen nem egy belsőből lépünk ki egy külsőhöz, hanem belátjuk, hogy a tudat eleve csak tárgyutdatként lehetséges, vagyis a tudat már eleve és közvetlenül a tárgynál van (Ullmann 2006. 244–245).

Azért vagyunk elégedetlenek, mert ha figyelmesen olvassuk a szöveget, akkor vagy azt találjuk benne, hogy *mi nem* az intencionalitás, vagy pedig annak folytonos ismétlését, hogy a tudat eleve a tárgynál van. Ullmann egész egyszerűen nem oldja meg az intencionalitás (egyébként Husserl által megfogalmazott) alapproblémáját, jelesül azt, hogy *miként képes egy tudati aktus vagy tudatállapot egy tőle különböző tárgyra irányulni?* E kérdésre nem válasz az, ha leszögezzük, hogy a tudat – fenomenológiai perspektívából nézve – „kint” van a tárgynál, ugyanis épp a tudat e „kintlévése” az a tény, *amelyet magyarázni kell.* Ullmann egész egyszerűen összekeveri a magyarázandót a magyarázattal, mégpedig úgy, hogy magyarázatnak tekinti a magyarázandót. Később így folytatja:

A fenomenológia [...] az újkori reprezentáció-elmélet egészét veti el, és azt állítja, hogy 1. az intencionalitás nem két különálló és a viszonytól függetlenül is létező entitás egymásra vonatkozása: tudati és tárgyi oldal csak az intencionalitás mozzanataiként értelmezhető. 2. A tudat nem egy belső képen keresztül vonatkozik a tárgyra (Ullmann 2006. 245).

Ullmann fenti állítását kétféleképpen tudjuk értelmezni. Egyrészt az intencionalitásról szóló 1. megállapítás a „tudat” és a „tárgy” szavak definícióit tartalmazza. Ha így áll a dolog, akkor Ullmann nyilván szükségszerű igazságnak tartja, hogy a tudat és a tárgy *kizárólag* a kettő között fennálló intencionális viszonyban létezik. Ennek a definíciónak (ha ugyan erről van itt szó) a jogosultságát valamivel igazolni kellene, ám erre Ullmann még csak kísérletet sem tesz. Másrészt az az állítás, hogy „a tudati és tárgyi oldal *csak* az intencionalitás mozzanataként értelmezhető”, nem szükségszerű igazság, hanem a világunk leírását nyújtja. Ebben az esetben egy kontingens állításról van szó, amely más lehetséges világokban nem feltétlenül igaz. Természetesen ahhoz is érvelésre, vagy legalábbis intuitív megfontolásokra lenne szükség, hogy a mi világunk ténylegesen ilyen, nem pedig olyan, amelyben az elme reprezentációkat alkot a (feltehetően) rajta kívül létező valóságról. Ullmann azonban nem ajánl ilyen érvet vagy megfontolást. (Ezzel szemben lásd Walker 1978. 110–118 érveit a tárgyfogalom nélküli tudatos tapasztalat lehetőségéről.)

4. A tanulmány *Jelentés és tárgyra irányulás* című része kapcsán nem érziünk kész-tetést az állásfoglalásra, ugyanakkor *A természettudományokhoz való viszony* című részhez van hozzáfűznivalónk. Ullmann itt így ír:

[az analitikus filozófusok] meggyőződés[e] szerint [...] a valóság végső szer-kezetét a természettudományok tárják fel [...] [s ]logikusan következik eb-ből, hogy akkor a jelentés alapfelfogása nem is lehet más, mint a referencia modell [...] [és, hogy] a tudatról [n]em érdemes túl sokat spekulálni, elég kivárni azt a néhány évet, amire a neurobiológia megadja a tudatnak neve-zett illúzió tudományos magyarázatát (Ullmann 2006. 246–247).

Erről csak annyit: e két sommás megállapítást korábban már cáfoltuk. Egyrészt a jelentés referenciaelméletének nincsen köze a fizikalista tárgyontológiához, másrészt a kortárs analitikus filozófusok igen nagy része foglalkozik a fenome-nális tudatosság problémájával.

Ami a fenomenológia és a természettudományok viszonyát illeti, Ullmann szerint a következő a helyes álláspont:

[A] fenomenológia a *tudományosság eszméjének* mint különös értelemalakzat-nak a természetét, struktúráját, alapmeggyőződéseit, történeti kialakulását stb. próbálja feltárni. Ebben a megközelítésben a tudomány 1. egy sajátos életvilág (az európai kultúra) és egy sajátos történeti kor (a modernitás) ideo-lógiájaként és vezéreszméjeként mutatkozik meg, másrészt pedig 2. a tuda-mányosság eszméje mögött uralomra jutott naturalista-objektivistá felfogás úgy jelenik meg, mint ami felelőssé tehető az élet túlzott technicizálódásá-ért és az ebből következő kulturális válságért (Ullmann 2006. 247).

Három rövid megjegyzés ezzel kapcsolatban. Egyrészt úgy gondoljuk, hogy az európai természettudományok, illetve a tudományos haladás relativizálóinak (így Ullmann-nak is) magyarázatot kellene adniuk a kortárs természettudomá-nyos elméletek interkulturális érvényességére. Továbbá – álláspontjukat alá-támasztandó – olyan szilárd alapot kellene találniuk, amelyen állva magyarázni tudják a tudományos haladás és ezen érvényesség állítólag illuzórikus voltát. Ilyenről azonban nem tudunk.

Másrészt a technicizálódással kapcsolatban úgy gondoljuk, hogy az ész (így a különféle technológiák alapjául szolgáló tudományos felfedezések is) önma-gában semmiféle kulturális válságért nem tehető felelőssé, mivel jó és rossz célokra egyaránt felhasználhatók. Nincs értelme tehát a tudományokat felelőssé tenni, vagy démonizálni, hiszen a felhasználóktól függ, milyen gyakorlati célok-ra alkalmazzák az elméleti eredményeket.

Harmadrészt annak latolgatását, hogy a technika vívmányai összességükben jók vagy inkább rosszak, a magunk részéről értelmetlennek tartjuk, és nem ér-

tünk egyet Michael Dummett-tel (1981), aki szerint jobb lett volna, ha a fizikai tudományok fejlődése megállt volna 1900-ban. Szerintünk sehová sem vezetne, ha – mondjuk – elkezdenénk összehasonlítani a gázkamrák és az atombombák áldozatainak számát a modern orvostudomány eszközeivel megmentett betegek számával. Ennek eldöntéséhez ugyanis a potenciális jövőbeli károkkal/jóvéte-ményekkel is számolni kellene, ezeket azonban senki sem láthatja előre.

5. A tanulmány *Fordulat* című részében a következő áll:

A fenomenológia és a rokon irányzatok alapkövetelménye a fordulat, vagyis a természetes beállítottság feladása és egy filozófiai, reflexív beállítottság felvétele. Az analitikus filozófia pedig éppen a természetes beállítottságról nem hajlandó lemondani, ebből következik az önmagára vonatkozó reflexió hiánya, az egész helyett a részproblémák vizsgálata, a gondolkodás technicizálása. Mi a természetes beállítódás? A hétköznapi józan ész, de különösen a tudomány meggyőződéseinek reflektálatlan elfogadása. És mi a fordulat? Annak belátása, hogy a filozófia nem is lehet más, mint a hétköznapi és tudományos józan ész igazságainak feladása (Ullmann 2006. 248).

Ahogy korábban *Mi és mi nem az analitikus filozófia?* (2007) című cikkünkben bemutattuk: (a) az analitikus filozófia ipari méretekben vizsgálja saját előfeltevéseit, (b) tisztában van a különböző részproblémák egymáshoz való viszonyával, és (c) igen nagyszámú az olyan analitikus művek száma, amelyek az „egész-szel” foglalkoznak.

Ha Ullmann komolyan gondolja azt, hogy „a filozófia *nem is lehet más*, mint a hétköznapi és a tudományos józan ész igazságainak *feladása*” (Ullmann 2006. 248. Kiemelés tőlünk, E. B., T. J.), akkor úgy gondoljuk, értelmetlen bármiféle párbeszéd. E kijelentés ugyanis semmi más, mint azoknak a filozófusoknak a kirekesztése, akik nem hajlandók feladni a „tudományos józan ész-t”. Ugyan-ebben a részben Ullmann még a következőket is állítja:

Az analitikus filozófia nem csupán a fenomenológiától és a kontinentális filozófiáktól idegenkedik, hanem minden olyan gondolkodótól, aki valamilyen értelemben radikális fordulatot követelt a gondolkodásban. Két szimbolikus figura merül fel ebben az összefüggésben, Kant és Wittgenstein. Az analitikus filozófia mindkettőre legalább annyi gyanakvással tekint, mint a klasszikus, XX. századi kontinentális gondolkodókra. Mindkettő radikális fordulatot követelt a gondolkodástól és olyan meggyőzésekkel szállt szembe, amelyek mélyen meghatározzák az analitikus világgépet (Ullmann 2006. 248).

Még mielőtt belemennénk a részletekbe: nem igaz, hogy a kortárs analitikus filozófusok „hasonló gyanakvással” tekintenek Kantra és Wittgensteinre, mint a kortárs kontinentális filozófusokra. Mégpedig azért nem, mert sem Kantra, sem (a kései) Wittgensteinre nem tekint gyanakvással az analitikus szakma, még akkor sem, ha időnként keményen bírálja is mindkettejük nézeteit.

Továbbá: súlyos tévedés azt gondolni, hogy Wittgenstein nézeteivel azért nem rokonszenvez a kortárs analitikus filozófusok egy része, mert Wittgenstein amolyan „kellemetlen figura”, aki mindig saját előfeltevéseinkkel szembesít. Akik elutasítják Wittgenstein nézeteit, azok azért utasítják el, mert szerintük e nézetek nagy része nem állta ki a kritikai vizsgálat próbáját. Kantról Ullmann a következőképpen vélekedik:

A kontinentális filozófia sok tekintetben Kant örököse, de nem a naiv módon félreértelmezhető idealizmus, hanem a fordulat örököse. A tekintet átállításának és annak a mély belátásnak az örököse, hogy a dolgok nem léteznek magukban, a tudattól függetlenül, ahogy azt általában, reflektálatlanul hisszük (Ullmann 2006. 249).

Engedtessek meg nekünk, hogy eltekintsünk Kant analitikus értelmezőinek felsorolásától, mert a lista igen hosszúra nyúlna. Azt viszont megjegyeznénk: nemcsak a kontinentális, hanem az analitikus filozófusok közül is sokan Kant örökösének tartják magukat. Ahogy például Robert Hanna fogalmaz: „Az analitikus filozófia története Fregétől Quine-ig az analicitásfogalom felemelkedésének és bukásának története, amelynek eredetét és jellemzőit egyaránt Kant első *Kritikája* tartalmazza” (Hanna 2001. 121). Searle pedig így ír: „az analitikus filozófiát tekinthetjük Locke, Berkeley, Hume [...] empirizmusa és Kant transzcendentális filozófiája természetes leszármazottjának” (Searle 2003. 2).

Ami a dolgok tudatfüggetlenségét illeti: a „tekintet” Ullmann által javasolt átállítással az egész kortárs realizmus–antirealizmus-vita értelmetlenné válna. Hiszen a vitatott kérdés pont az: létezhetnek-e dolgok bármilyen rájuk vonatkozó gondolattól, ismerettől, tapasztalattól, illetve beszédétől függetlenül.

Végezetül nem tudjuk elhallgatni: szerintünk a kései Wittgenstein híres privátnyelv-érve alapjaiban szembenáll a karteziánus meggyőződésen alapuló fenomenológia programjával, amely szerint léteznek lényegileg privát, csupán egyetlen személy által megfigyelhető, és csak a beszélő privát nyelvén leírható tények. Csak éppen az Ullmann által emlegetett Wittgenstein fogalmazta meg ezt az érvet kései időszakában.

## III. AZ EGYESÜLÉS EGY FENOMENOLÓGUS SZEMÉVEL

Ullmann véleménye szerint:

A filozófia két fő irányzata újra egyesülni fog, éppen azért, mert konkrét kulturális és életproblémákkal kell szembenézniük: az analitikus filozófia fel fogja adni merev naturalizmusát és skolasztikus életidegenségét, a „kontinentális irányzatok” pedig a többszörösen metafizikai spekulációkból kénytelenek visszatérni a tárgyyszerű és konkrét vizsgálódásokhoz (Ullmann 2006. 251).

Tegyük fel, hogy igaz Ullmann jövendölése. De hogyan képzei e „schilleri egyesülés” munkamegosztását?

Az értelem szubjektív vagy fenomenális aspektusát a fenomenológia, hermeneutika stb. fogja láthatóvá tenni. A szubjektív aspektus az értelem tudati oldala: az, ami – éppen a tudattal való kapcsolata miatt – nem mérhető ki teljesen, amelyben új és új rétegek tárulnak fel szakadatlanul, ami új és új nézőpontokból mutatkozik meg, ami sohasem rögzíthető véglegesen. Az objektív aspektus pedig az analitikus hagyomány felől fog jobban látszani. Az objektív aspektus az értelem mindenkor „rögzített” oldala: az, hogy mégiscsak közösen értett, vagyis mások által is átélt, elgondolt tapasztalatok között élünk, ezeknek van rögzíthető, informatív tartalmuk, logikai vázok, megjósolható lefolyásuk és modellezhető szerkezetük (Ullmann 2006. 251).

Úgy tűnik, Ullmann úgy képzei ezt az egyesülést, hogy míg a fenomenológia (és a hermeneutika) az új gondolatokat szállítja, addig az analitikus filozófia majd elemzi és szisztematikus formába öltözteti e már „rögzített” gondolatokat. Vagyis a kontinentális filozófusok végzik a kreatív munkát, az analitikusok pedig a mechanikusát. E javaslat háttérében nyilván Ullmann azon meggyőződése áll, hogy míg a fenomenológusok (és a hermeneuták) tobzódhatnak az új ötletekben, addig az analitikus kollégák amolyan ötlettelen savanyújancsik, akik csak a már „rögzített” gondolatok elemzésére képesek.

Jól látható mindez abból, hogy Ullmann milyen fogalmakat asszociál a kontinentális filozófiához és milyeneket az analitikus filozófiához. A kontinentálishoz a következőket: „új és új rétegek szakadatlan feltárulása”, „új és új nézőpontok”, „végleges rögzíthetetlenség”. Az analitikushoz pedig: „rögzítettség”, „logikai váz”, „megjósolható lefolyás”, „modellezhetőség”.

A bevezetésben is megemlíttük, hogy e vizionált munkamegosztást egyrészt életszerűtlennek, másrészt ésszerűtlennek tartjuk.

Azért életszerűtlen, mert erősen kételkedünk abban, hogy a kontinentális filozófia képviselői új ötletekben tobzódó, „új és új rétegeket szakadatlanul fel-

táró” figurák volnának. Többségük inkább – legalábbis a szemünkben – egy-egy „nagy” kontinentális filozófus nagysága előtt leboruló, annak szövegeit elemző, a valódi filozófiai problémákat messze elkerülő alak. Ahogyan azt a *Kellék* című folyóiratban írtuk: „a legtöbb fenomenológus Husserlről vagy valamely társáról ír, és *nem csinálja* a fenomenológiát” (Eszes–Tözsér 2005. 67, kiemelés az eredetiben). Nem kérünk mást, mint hogy az elfogulatlan olvasó nézze meg e két iskolához tartozó szerzők tanulmányait.

Azért ésszerűtlen, mert az analitikus filozófusok is igen komolyan foglalkoznak az egyes szám első személyű (szubjektív) perspektívával. E sorok egyik írója (T. J.) épp egy olyan észleléselemélet kidolgozásán munkálkodik, amely a lehető legjobban illeszkedik az érzéki tapasztalat fenomenológiájához. És ebben a vállalkozásban – finoman fogalmazva – nincsen egyedül.

## IRODALOM

- Allaire, Edwin B. 1963/1998. Bare Particulars. In Stephen Laurence – Cynthia Macdonald (szerk.) *Contemporary Readings in the Foundations of Metaphysics*. Oxford, Blackwell. 248–254.
- Allaire, Edwin B. 1965/1998. Another Look at Bare Particulars. In Stephen Laurence – Cynthia Macdonald (szerk.) *Contemporary Readings in the Foundations of Metaphysics*. Oxford, Blackwell. 259–263.
- Armstrong, David M. 1968. *A Materialist Theory of the Mind*. London, Routledge.
- Armstrong, David M. 1980. Identity through Time. In Peter van Inwagen (szerk.) *Time and Cause: Essays in Honor of Richard Taylor*. Dordrecht, Reidel. 67–78.
- Benatar, David 2006. *Better Never to Have Been*. Oxford, Clarendon Press.
- Bergmann, Gustav 1967. *Realism*. Madison/WI, University of Wisconsin Press.
- Beyerstein, Barry L. 1991. The Brain and Consciousness: Implications for Psi Phenomena. In Kendrick Frazier (szerk.) *The Hundredth Monkey*. Buffalo/NY, Prometheus Books. 43–53.
- Bostrom, Nick 2003. Are You Living in a Computer Simulation? *Philosophical Quarterly*. 53. évf. 211. sz. 243–255.
- Campbell, Keith 1981/1998. The Metaphysics of Abstract Particulars. In Stephen Laurence – Cynthia Macdonald (szerk.) *Contemporary Readings in the Foundations of Metaphysics*. Oxford, Blackwell. 351–363.
- Campbell, Keith 1988. Philosophy and Common Sense. *Philosophy*. 63. évf. 244. sz. 161–174.
- Campbell, Keith 1990. *Abstract Particulars*. Oxford, Blackwell.
- Casullo, Albert 1984. The Contingent Identity of Particulars and Universals. *Mind*. 93. évf. 372. sz. 527–541.
- Casullo, Albert 1988. A Fourth Version of the Bundle Theory. *Philosophical Studies*. 54/1. 125–139.
- Chalmers, David 2006. The Representational Character of Experience. In Brian Leiter (szerk.) *The Future of Philosophy*. Oxford, Oxford University Press. 153–181.
- Churchland, Paul M. 1981/1991. Eliminative Materialism and the Propositional Attitudes. In David M. Rosenthal (szerk.) *The Nature of Mind*. New York – Oxford, Oxford University Press. 601–612.
- Churchland, Patricia S. 1986. *Neurophilosophy*. Cambridge/MA, MIT Press.

- Cleve, James van 1985/1998. Three Versions of the Bundle Theory. In Stephen Laurence – Cynthia Macdonald (szerk.) *Contemporary Readings in the Foundations of Metaphysics*. Oxford, Blackwell. 264–274.
- Crane, Tim 2001. *Elements of Mind*. Oxford, Oxford University Press.
- Dennett, Daniel C. 1971/1998. Az intencionális rendszerek. In uő *Az intencionalitás filozófiája*. Ford. Pléh Csaba. Budapest, Osiris. 7–38.
- Dummett, Michael 1991/2000. *A metafizika logikai alapjai*. Ford. Szalai Miklós et al. [a tanulmányban hivatkozott fejezet Csaba Ferenc fordítása]. Budapest, Osiris.
- Dummett, Michael 1981. Ought Research to Be Restricted? *Grazer Philosophische Studien*. 12–13. sz. 281–298.
- Eszes Boldizsár – Tözsér János 2005. Mi az analitikus filozófia? *Kellék*. 27–28. sz. 45–71.
- Eszes Boldizsár – Tözsér János. 2007. Mi és mi nem az analitikus filozófia? *Magyar Filozófia* *Szemle*. 51/1–2. 75–108.
- Feyerabend, Paul 1963. Comment: Mental Events and the Brain. *Journal of Philosophy*. 60/11. 295–296.
- Frege, Gottlob 1918/2000. Logikai vizsgálódások I–II. rész. Ford. Máté András. In Frege, Gottlob *Logikai vizsgálódások*. Budapest, Osiris. 191–239.
- Fodor, Jerry 1987. *Psychosemantics*. Cambridge/MA, MIT Press/Bradford Books.
- Gallagher, Shaun – Zahavi, Dan 2008. *The Phenomenological Mind. An Introduction to Philosophy of Mind and Cognitive Science*. London, Routledge.
- Hanna, Robert 2001. *Kant and the Foundations of Analytic Philosophy*. Oxford, Clarendon Press.
- Heller, Mark 1990. *The Ontology of Physical Objects: Four Dimensional Hunks of Matter*. Cambridge, Cambridge University Press.
- Heller, Mark 1993. Varieties of Four Dimensionalism. *Australasian Journal of Philosophy*. 71/1. 47–59.
- Higginbotham, James 2006. Truth and Reference as the Basis of Meaning. In Michael Devitt – Richard Hanley (szerk.) *The Blackwell Guide to the Philosophy of Language*. Oxford, Blackwell. 58–76.
- Hochberg, H. 1964. Things and Qualities. In W. Capitan – D. Merrill (szerk.) *Metaphysics and Explanation*. Pittsburgh, University of Pittsburgh Press. 123–160.
- Kálmán László – Rádai Gábor 2001. *Dinamikus szemantika*. Budapest, Osiris.
- Kripke, Saul 1972/2007. *Megnevezés és szükségszerűség*. Szerk., bevez. Zvolenszky Zsófia; ford. Bárány Tibor. Budapest, Akadémiai Kiadó.
- Lamont, Corliss 1990. *The Illusion of Immortality*. New York, Unger/Continuum.
- Lewis, David 1966. An Argument for the Identity Theory. *Journal of Philosophy*. 63/1. 17–25.
- Lewis, David 1986. *On the Plurality of Worlds*. Oxford, Basil Blackwell.
- Lewis, David 2001. Truthmaking and Difference-Making. *Noûs*. 35/4. 602–615.
- Lewis, David 2002. Tensing the Copula. *Mind*. 111. évf. 441. sz. 1–14.
- Martin, Charles Burton 1980. Substance Substantiated. *Australasian Journal of Philosophy* 58/1. 3–10.
- McGinn, Colin 1991. *The Problem of Consciousness*. Oxford, Blackwell.
- Millikan, Ruth Garrett 1984. *Language, Thought and Other Biological Categories*. Cambridge/MA, MIT Press.
- Nagel, Thomas 1974/1997. What It Is Like To Be a Bat? In Ned Block – Owen Flanagan – Güven Güzeldere (szerk.) *The Nature of Consciousness*. Cambridge/MA, MIT Press. 519–527.
- Putnam, Hilary 1981. *Reason, Truth and History*. Cambridge, Cambridge University Press.
- Quine, Willard Van Orman 1960. *Word and Object*. Cambridge/MA, MIT Press.

- Quine, Willard Van Orman 1963. Identity, Ostension, and Hypostasis. In uő *From a Logical Point of View*. Evanston, ILL., Harper and Row. 65–79.
- Quine, Willard Van Orman 1976. Whither Physical Objects. In Robert Sonné Cohen – Paul K. Feyerabend – Marx W. Wartofsky (szerk.) *Essays in Memory of Imre Lakatos*. Dordrecht, Reidel. 497–504.
- Quine, Willard Van Orman 1981. *Theories and Things*. Cambridge/MA, Harvard University Press.
- Ramsey, William 1991. Where Does the Self-Refutation Objection Take Us? *Inquiry*. 33/4. 453–465.
- Rey, Georges 1983. A Reason for Doubting the Existence of Consciousness. In Richard J. Davidson – Gary E. Schwartz – David Shapiro (szerk.) *Consciousness and Self-Regulation* 3. köt. New York, Plenum. 1–39.
- Rey, Georges 1988. A Question About Consciousness. In Herbert R. Otto – James A. Tuedio (szerk.) *Perspectives on Mind*. Dordrecht, Reidel. 5–24.
- Rorty, Richard 1965. Mind–Body Identity, Privacy and Categories. *Review of Metaphysics*. 19/1. 24–54.
- Rorty, Richard 1970. In Defense of Eliminative Materialism. *Review of Metaphysics*. 24/1. 112–121.
- Rorty, Richard 1979. *Philosophy and the Mirror of Nature*. Princeton, Princeton University Press.
- Searle, John 2003. Contemporary Philosophy in the United States. In Nicholas Bunnin – Eric P. Tsui-James (szerk.) *The Blackwell Companion to Philosophy*. Oxford, Blackwell. 1–22.
- Sider, Theodore 1996. All the World's a Stage. *Australasian Journal of Philosophy*. 74/3. 433–453.
- Sider, Theodore 2001. *Four-Dimensionalism*. Oxford, Oxford University Press.
- Simons, Peter 1994/1998. Particulars in Particular Clothing: Three Trope Theories of Substance. In Stephen Laurence – Cynthia Macdonald (szerk.) *Contemporary Readings in the Foundations of Metaphysics*. Oxford, Blackwell. 364–384.
- Stich, Stephen P. 1983. *From Folk Psychology to Cognitive Science*. Cambridge/MA, MIT Press.
- Strawson, Peter Frederick 1959. *Individuals: An Essay in Descriptive Metaphysics*. London, Methuen.
- Szigeti Artilla 2002. A tudatosság: rejtély vagy képesség? Az analitikus filozófia és a kognitív tudomány találkozása a fenomenológiával. *Kellék*. 22. 81–122.
- Ullmann Tamás 2006. Analitikus és kontinentális filozófia. *Magyar Filozófiai Szemle*. 50/3–4. 227–251.
- Walker, Ralph C. S. 1978. *Kant*. London, Routledge.
- Williams, Donald C. 1953/2004. The Elements of Being. In Tim Crane – Katalin Farkas (szerk.) *Metaphysics: A Guide and Anthology*. Oxford, Oxford University Press. 262–272.



TÓTH ZITA VERONIKA

## Borbély Gábor és a középkori filozófia

Borbély Gábor, *Civakodó angyalok*. Budapest, Akadémiai Kiadó.  
2008. 379 oldal

Borbély Gábor könyve tagadhatatlanul hiánypótló, hiszen az első magyar nyelven írott szakkönyv, amely bevezetést nyújt a középkor filozófiájába. A könyv tehát egyrészt *bevezetés*, azonban – amint címe és előszava is utal rá – nem kíván szisztematikus lenni abban az értelemben, hogy a középkori filozófia minden jelentős területét tárgyalná. Így, míg gazdag képet nyújt a logika, a nyelv és az ismeretelmélet körébe tartozó vitákról, szinte egyáltalán nem tárgyal természet- vagy társadalomfilozófiai kérdéseket, és csak korlátozottan érint bizonyos etikai problémákat.

A bevezetés célközönsége, bár ezt Borbély Gábor külön nem jelzi, a filozófiában már jártas olvasó. A fejezetek ugyan ritkán támaszkodnak filozófiatörténeti előismertekre – a szükséges előzményeket a szerző általában kimerítően tárgyalja –, néhány érvmenet megértése mégis feltételezi az analitikus filozófia eszköztárának ismeretét. Ez azonban nem véletlen. A könyvnek ugyanis amellet, hogy megismertesse az olvasót a középkori filozófiával, egy további bevallott célja is van, mégpedig e filozófia „kifinomult logikai apparátusának” (9) bemutatása.

Másrészt Borbély Gábor könyve valóban az első magyar mű, amely átfogóan tárgyalja a középkori filozófiát, és ez egyúttal azt a kellemetlen tényt is jelenti, hogy mind ez ideig nem volt a középkori

filozófiának egységesen elfogadott magyar szaknyelve. Így, hogy csak a legalapvetőbb hiányt említsük, a középkori nevek átírásának sincs, egy-két népszerű szerzőt leszámítva, bevett módja; bár Aquinói Tamás és Canterburyi Anzelmet általában magyarosítva szoktuk emlegetni, ez korántsincs így például William Ockham vagy Petrus Aurcoli esetében. Borbély Gábor az egységesítés szakszerű, ám néha kényelmetlen módját választotta, amennyiben minden nevet, legalábbis első előfordulásakor, latinul említ. Bár e megoldás kétségtelen előnye, hogy egységessé teszi a neveket anélkül, hogy nem létező magyar fordításokba kényszerítené a szerzőt, ugyanakkor azzal a furesasággal is jár, hogy a megszokott Canterburyi Anzelmből Anselmus Cantuariensis, Adam Wodehamból Adamus Goddamus, vagy Walter Burley-ből Gualterus Burlaeus lett.

A tárgyalás középpontjában a középkor néhány filozófiai vitája áll – ez magyarázhatja a könyv címének egy részét, a „civakodó” kifejezést –, ezzel szemben némileg meglepő, hogy felépítése mégis inkább kronologikus. Noha a kötet egy-egy problémához kapcsolódóan gyakran áttekinti annak egész történetét, s így fejezetei túlnyúlnak az egyes szerzők által felölelt időszakokon, a vitákat azok központi szereplői köré csoportosítva mutatja be Abaelardustól a 14. századi buridanizmusig huszon-

egy fejezeten keresztül. (Bár ez a könyvből nem teljesen világos, valószínűleg e szereplőkre utalnak a címbeli „angyalok”, ha őket mint tiszta intelligenciákat értjük.) A részletesen bemutatott gondolkodók a következők: Petrus Abaelardus, Canterburyi Anzelm, Albertus Magnus, Aquinói Tamás, Bonaventura, Duns Scotus, Petrus Aureoli, Richard Campsall, William Ockham, Walter Burley, Walter Chatton, Adam Wodeham, Robert Holcot s végül Jean Buridan. A szerzők bemutatása rövid életrajzzal kezdődik, amelyet néhány jellemző tanításuk vagy érvek rekonstrukciója követ. Ez utóbbi szükségképpen szelektív, s Borbély Gábor könyvében, mint már jeleztük, e jellemző tanítások általában a logika, nyelvfilozófia, illetve episztemológia körébe tartoznak.

Így Canterbury Anzelm esetében, noha kétségkívül legnagyobb hatást gyakorolt érve, az ontológiai istenérv is bemutatásra kerül, a legnagyobb hangsúlyt az isteni előretudás és a szabad akarat összeczegyztetetőségének problémája kapja, amelyet a fejezet az arisztotelészi előzményektől kezdve Ágostonon és Boëthiuson keresztül részletesen tárgyal. Arisztotelész fatalista érve és különböző interpretációi meglehetősen technikai nyelvezetet kívánnak, azonban az alapvető fogalmakat – mint a *necessitas consequentis*, illetve *consequentiae*, a *sensus compositus* és *divisus*, vagy a modalitás statisztikus értelme – a fejezet kellőképpen tisztázza. Az elemzés megmutatja, hogy Anzelm megoldása Boëthiusra támaszkodik, amennyiben az örökkévaló jelen fogalmán alapul, és az „előretudás”, „eleve elrendelés” terminusokat az isteni örökkévalóság természetét figyelmen kívül hagyó metaforáknak tekinti.

Hasonlóképpen Petrus Abaelardus bemutatása során sem leginkább közismert etikája kerül középpontba, hanem az univerzálé-problémában tanítómesterével, Champeaux-i Vilmossal szemben elfoglalt

álláspontja. Bár az ábrázolás hitelét némiképp rontja az, hogy Abaelardus ellenfelei már a fejezet elején úgy jelennek meg, mint „manipulátorok” (74) és „machinátorok” (78, ez utóbbi jelzót Clairvaux-i Bernát kapta), az univerzálé-probléma bemutatása kimerítőnek mondható. A vita fő kérdése az univerzális nevek szignifikációjára vonatkozik, és annak ellenére, hogy a fejezet ismét kellőképpen tisztázza a használt fogalmakat – mint megnevezés, szignifikáció, *status*, *dictum*, referenciális kontextus stb. –, Abaelardus álláspontjának megértéséhez szükségeltetik némi nyelvfilozófiában való jártasság. Megoldása szerint ugyanis maga a kérdés rossz, amennyiben valóságos tárgyakká próbálja tenni a mentális objektumokat, és figyelmen kívül hagyja, hogy referenciálisan homályos kontextusban igaz állítást akkor is tehetünk valamiről, ha az csupán mint a gondolkodás belső tárgya létezik.

A 12. századi logikai iskolák (*logica vetus*, *logica nova*, realisták és nominalisták), az ugyancsak 12. századtól domináns eukleidészi tudományelmélet, valamint az ebben és az ezt követő században elterjedt új Arisztotelész-fordítások és ezek recepciójának rövid bemutatása után az Albertus Magnusra szánt csupán három oldalas fejezet némi hiányérzetet kelthet. Bár említésre kerül a filozófiai vizsgálódás autonómiájáról vallott nézete, egyáltalán nem esik szó logikájáról, metafizikájáról vagy lélekképzéséről, amelyek a nyugati Arisztotelész-recepció kezdetét jelentik.

A következő részletesen tárgyalt gondolkodó Aquinói Tamás. A létező és lényeg megkülönböztetésének bemutatása a *principium individuationis* és a lényeg megismerhetőségének problematikáját is érinti, és elsősorban a tamási megoldás nehézségeire hívja fel a figyelmet. Az „öt út” elemzése az előzőkhez képest meglehetősen rövid, igaz azonban, hogy ebben a tárgyban az olvasó bátran fordulhat Klima Gyula korábban

megjelent kiváló cikkéhez (Klima Gyula 1981. Az öt út – Aquinói Tamás istenbizonyítékai. *Világosság*, 22/12 Melléklet. 3–30). Úgyancsak futólag esik szó Tamás nagy *summáiról*, elsősorban a *Summa contra gentiles*ről, és összességében a fejezet nem tanúskodik arról, hogy Tamás – például lélefilozófiájával, ismeretelméletével, vagy a világ örökévalóságáról vallott nézeteivel – legalább olyan érdekes alakja a középkori vitaközás művelészetének, mint a főttebb bemutatott Abaelardus. (E hiányt némiképpen pótolja a következő fejezet, amely Bonaventura és Tamás vitáját elemzi.)

Az 1277-es, meglehetősen vitatott határási párizsi elítélő határozat, az ezzel kapcsolatba hozható Brabanti Siger és Boëthius de Dacia, valamint néhány kevésbé ismert 13. századi szerző rövid áttekintése után Johannes Duns Scotus részletesebb elemzése következik. Duns Scotus igazi esemege az analitikus filozófiát kedvelők számára, és ezt jól tükrözi a könyv e fejezete. Betekintést nyerhetünk Scotus szokatlan univocitás-elméletébe, amely szerint a transzcendentálekát egyazon értelemben állíthatjuk Istenről és minden dologról; az intuitív és absztraktív megismerés fogalmának és az ezzel kapcsolatos, a 13. század végétől kezdve egyre erősebb, jellegzetesen középkori szkeptikus érvek történetébe, amelyek megelőlegezik Descartes módszertani szépségszót; a *principium individuationis* problémájával kapcsolatos scotusi megoldásba, amely a sem az anyaggal, sem a formával, sem a kettőjük összetételével meg nem egyező *haecceitas* segítségével magyarázza az individualitást; abba a lélefilozófiai irányzatba, amely – Aquinói Tamással szemben – az akarat és értelem csupán formális különbözőségét állítja; s végül, de nem utolsósorban a szabad akarat és kontingencia szinkronikus értelmezésébe, amely szerint az akarat szabadsága a szinkronikusan alternatív lehetőségek közötti választásban áll.

Petrus Aureoli és Richard Campsall fontosabb tanításainak vázolója után egy újabb izgalmas gondolkodó, William Ockham következik, aki a könyv leghosszabb fejezetét mondhatja magáénak. Ez a könyv irányultságát figyelembe véve érthető, hiszen olyan logikai és nyelvelméleti fogalmak tárgyalására kerülhet itt sor, mint a terminusok sajátosságai (*proprietas terminorum*), a szignifikáció, a szuppozíció, az abszolút és konnotatív terminusok megkülönböztetése, valamint a mentális nyelv problémája. A fejezet szintén kitér a 20. században nagy figyelmet kapott ockhami ismeretelméletre és annak esetleges szkeptikus konzekvenciáira, amelyek különösen jól megfigyelhetők Utricurria idézett szövegében.

Az Ockham körüli és utáni logikai, szemantikai és ismeretelméleti vitákat négy szerző – Walter Burley, Walter Chatton, Adam Wodeham és Robert Holcot – illeszttrálja. Szó esik a modista grammatika és logika néhány párizsi képviselőjéről, valamint az oxfordi kalkulátorokról, akiknek természetfilozófiáját Borbély Gábor – igaz, a John Murdochtól származó, de vitatható jellemzést használva – egyszerűen csak „természet nélküli természetfilozófiaként” (328) írja le. Bár a könyv sehol nem fektet nagy hangsúlyt a természetfilozófiára, az oxfordi kalkulátorok mozgással kapcsolatos elmélete is jó példát jelenthetett volna a középkor kifinomult érvelési módszerének demonstrálásához.

A könyv Johannes Buridanusszal és tanítványaiával zárul. Bár Buridanus az utóbbi évtizedek egyik legnépszerűbb középkori gondolkodója, itt csupán szemantikájának alapjairól esik szó. Szuppozíció-elmélete, amely a Petrus Hispanustól származó természetes szuppozíció segítségével magyarázza a tudomány szükségyszerű állításainak lehetőségét (*nota bene*, ez nem az egyetlen szükséges feltétele a tudományos állításnak; az ennél valamelyest bonyolultabb

esszenciális predikáció kérdésével Borbély Gábor nem foglalkozik), valamint a Buri-danust követő évszázad reakciója illő befejezése a könyvnek: nem szűkülökődik civakodásban, és kiválóan megfelel a könyv logikai-szemantikai irányultságának.

A könyv tehát valóban bemutatja a 12. és 14. század közötti időszak legfontosabb filozófiai problémáit és filozófiai módszereit. Egy bevezetés természetesen soha nem lehet kimerítő; bár a könyv nem foglalkozik a középkori filozófia minden területével, és nem tárgyal olyan jelentős szerzőket, mint Eckhart mester vagy a későbbi Nicolaus Cusanus, témáját szakszerűen és a legújabb interpretációk ismeretében mutatja be. Fejezeteit az itt vázoltnál élvezhetőbbé teszik azok a marginálisok, amelyek – a középkori kommentárirodalom hagyományának megfelelően – Borbély Gábor könyvének is szerves részét képezik, és amellet, hogy áttekintést adnak a főszöveg témájáról, már-már túlzott könnyedségükkel egyszerűbbé, de legalábbis szórakoztatóbbá teszik annak megértését. A könyv egy további pozitív, a megértést nagymértékben segítő eleme a számos elsődleges szövegrészlet, amelyek a fejeze-

tek közben, de azoktól formailag elkülönülten tesznek elérhetővé eddig magyarul nem olvasható írásokat. Szintén az olvasó tájékozódását szolgálja a terjedelmes bibliográfia, amely, noha könnyebben kezelhető lett volna témánként vagy szerzőnként, illetve elsődleges és másodlagos irodalom szerint csoportosítva, így is további iránymutatást ad.

Sikerült-e, összességében, Borbély Gábornak bevezetnie az olvasót a középkori filozófiába, és meggyőznie arról, hogy e filozófia kifinomult logikai apparátussal rendelkezett problémái megoldásához? A fentebbiek fényében elmondhatjuk, hogy e célt a könyv elérte. Hozzátehetjük, hogy elérhette volna kevesebbrel is; valószínűleg Richard Campsall vagy Robert Holcot bemutatása nem elengedhetetlen tartozéka egy bevezető írásnak, ahogyan az érvek struktúrája is világosan követhető lett volna a formális logika apparátusa nélkül. Ezek azonban remélhetőleg nem fogják elijeszteni az olvasót e bevezetéstől, amely hasznos segédanyagul szolgálhat mind az egyetemi filozófiaoktatásban, mind a szélesebb, filozófiában művelt olvasóközön-ség számára.

## E számunk szerzői

- BALOGH LÁSZLÓ LEVENTE (1969) Debreceni Egyetem BTK Politikatudományi Tanszék, kutatási területe a politikatudomány és a politikafilozófia, e-mail: baldro@freemail.hu
- ESZES BOLDIZSÁR (1973) több analitikus filozófiai tanulmánykötet fordítója, kutatási területei a szemantika és a nyelvfilozófia, e-mail: eszes.b@pannonmail.hu
- GULYÁS PÉTER (1984) az Eötvös Loránd Tudományegyetem BTK Filozófia Doktoriskolájának doktorandusa, kutatási területe a karteizianizmus, e-mail: gp84@freemail.hu
- JOÓ MÁRIA, Eötvös Loránd Tudományegyetem BTK Filozófia Intézet, kutatási területe Platón, a filozófiai antropológia és a feminista elmélet (Simone de Beauvoir), e-mail: joo\_maria@ludens.elte.hu
- KICSÁK LÓRÁNT (1971) Eger, Eszterházy Károly Főiskola, Filozófia Tanszék, kutatási területe a német és francia fenomenológia, e-mail: kicsak@ektf.hu
- MOLDVAY TAMÁS (1968) Gábor Dénes Főiskola, Gazdasági és Társadalomtudományi Intézet, kutatási területe az újkori filozófia, e-mail: tamasmoldvay@yahoo.com
- SUTYÁK TIBOR (1968) Szegedi Tudományegyetem BTK, Filozófia Tanszék, kutatási területe a pszichológia filozófiája, legutóbbi könyve: *Michel Foucault gondolkodása*. Budapest, Attraktor, 2007, e-mail: sutyaktibor@gmail.com
- TAKÁCS ÁDÁM (1970) Eötvös Loránd Tudományegyetem BTK, Atelier Európai Társadalomtudomány és Historiográfia Tanszék, kutatási területei a történelemfilozófia, a klasszikus és kortárs fenomenológia és a 20 századi francia filozófia, e-mail: takacsadam@hotmail.com
- TÖZSÉR JÁNOS (1970) MTA-ELTE Nyelvfilozófiai Kutatócsoport, kutatási területei: elmefilozófia, metafizika, észleléselemélet, legutóbbi könyve: *Metafizika*. Budapest, Akadémiai Kiadó, 2009, e-mail: jantozser@gmail.com

---

## Summaries

### The Body–Body Problem

TIBOR SUTYÁK

In recent philosophical debates there have been two standard ways of approaching the concept of the body. Analytic philosophers of mind, theorizing over the problems of qualia and mental causation, provide us with a genuine philosophical puzzle: promising candidates for solving the former problem turn the latter problem inextricable, and *vice versa*. However, in describing this puzzle, philosophers of mind operate with a rather reductive notion of the body. “Body” means for them “brain” or some kind of neuro-chemical agent. Phenomenologists present us with a much more sophisticated concept of the body, which includes and differentiates between the physico-biological and the experiential or „lived” aspect of the former. This twofold body-concept generates a new problem, the so-called body–body problem, which concerns the relationship between the physical and the phenomenological aspects of one’s own body.

In the paper I provide a brief overview of the idea of the “embodied mind”. In this reconstruction I focus on the critique of representation, the principle of situatedness,

the role of explorative motion and the importance of affectivity. Furthermore — after proposing a short argument against the possibility of the existence of philosophical zombies — I try to apply the theoretical tools invented by the proponents of the “embodied mind” theory to deal with the puzzle of the mind–body problem which now emerges as the “body–body problem”.

### **Transcendentalism and the Own Body in Husserlian Phenomenology** LÓRÁNT KICSÁK

The main aim of the paper is to contextualize the problem of the body in Husserlian phenomenology, and to show how phenomenological insights concerning our own-body, being labelled as “novelties” by Husserl, modified several early principles of phenomenology. I present how corporeality became the vehicle of the plan of phenomenology as a universal science, and explore the conflicts emerging between the plan of a universal science and the insights of the own-body analyses. In my view, the own-body analyses fundamentally undermine the possibility that phenomenology could become an established form of a final and universal science of the world, and not only a radical form of philosophers’ self-knowledge.

According to Husserl, if we transilluminate ourselves as transcendental intersubjective subjects, it is necessary to reach the science of universal reason and true knowledge as a result thereof. In order to achieve this, it should be possible to bring into a scientific relation *factum* and ideal possibility and necessity. As a consequence, Husserlian phenomenology, despite its original intention, became metaphysics, or at least it had to rely on a type of Leibnizian metaphysics.

### **“Subjective Objectivity”. Husserl’s Phenomenological Theory on the Foundational Role of the Body** ÁDÁM TAKÁCS

It is well-known that from the 1920’s Husserl connects more and more intimately the phenomenological analysis of consciousness with that of the world as the horizon of experience. However, it is not always made clear in what way this new approach modifies the doctrine of transcendental subjectivity. In this paper I will argue that in defining “concrete subjectivity” as a world-experiencing being, Husserl not only rediscovers the phenomenological significance of the body, but also endows it with a foundational character. I will try to show that this foundational structure has three representative forms. First, the body as feeling and as a capacity for movement contributes fundamentally to the constitution of the self-manifestation of concrete subjectivity. Second, it is the primary source of objectivity in so far as the body lies in the centre of the subject–object relation. Third, the body is the fundament of intersubjectivity, because it is always in its bodily form that one can encounter the other in the field of original experience.

### **Feminist Theory and the (Female) Body** MÁRIA JOÓ

The paper provides a short outline of corporeal feminism, presents its main theoretical positions and authors in order to promote the exchange between the main (male)

discourse in Hungarian philosophy and feminist theory. It concentrates on the debate between two major positions in gender-constructionism; one is more influenced by the phenomenology of Merleau-Ponty and his contemporary follower and friend Simone de Beauvoir, the other by post-structuralist philosophers, like Foucault. Both approaches agree that the body has an ambiguous – cultural as well as natural – existence. Of the many authors, the paper concentrates on the works of Iris Marion Young, Susan Bordo and Judith Butler. In the list of references the Hungarian secondary literature on feminist theory (mainly translations) is also presented.

### Spinozistic Body as a Problem

TAMÁS MOLDVAY

In this paper, I address the following question: what is the body in Spinoza's *Ethics*? In the first part I examine the "particular thing" and the "individual" as two Spinozistic concepts of the body, which do not appear to be equivalent with one another. I show that the concept of the body involves implicitly the perspective of an incorporating causality, that is, if several bodies concur in one action, this action could cause one particular body. After the analysis of the Spinozistic concept of "the mutual relations of motion and rest", in the second part of my paper, I emphasize the importance of the separation of the perspective of the "causa seu ratio" from the perspective of the incorporating causality. Finally, I show that the individuation inherent in incorporating causality can result in chimerical bodies.

### The Anatomy of Leviathan

LÁSZLÓ LEVENTE BALOGH

All forms of rule are linked to those visual and rhetoric representations or media through which they appear to the public. These representations are primarily attached to the institutions of rule, practices of violence and to the political community. Significantly, they always involve references to the imaginary aspects of politics. Visual and rhetoric representation not only convey the means and institutions of rule, but also constructs a sphere of politics that is imaginary in its essence. This sphere in turn determines the very nature of the rule. The ideal and probably most widely used medium of this aesthetic procedure is the human body which is abundant in metaphoric meanings. Out of all political bodies Hobbes' Leviathan made the largest impact. This image was capable of embracing all concepts associated with the imaginary political sphere and also includes mechanical, godlike, human and beastly features. In the background of all mentioned forms we find the principle of order in which means and concepts of order are inherently intertwined and institutionalised.

### Sympathie and "Corporeal Innatism" in Descartes

PÉTER GULYÁS

The physiological writings of Descartes are usually evaluated as marginal and insignificant parts of his philosophy, which lack any kind of interest for scholars in our days. In this paper I intend to demonstrate that the examination of these writings is indispensable, if we seek to understand Cartesianism in its historical context. My main aim is to show how the physiological explanation of the communication between the maternal imagination and

the foetal development gains metaphysical significance in Descartes' works. As a first step, I try to understand the mechanical and corporeal process of imagination described by Descartes in the *Traité de l'homme*. In the next section, the paper follows closely the fourth and fifth chapters of the *Description du corps humain* in order to outline the main characteristics of Cartesian embryology. In the third section I tackle the problems originating in Descartes' attempt to interpret the connection between the maternal brain and the foetal body in a mechanistic way. Finally, I will show that Descartes had strong metaphysical reasons to insist on the existence of *sympathie* between the mother and the embryo.

### Analytic Philosophy and Phenomenology BOLDIZSÁR ESZES – JÁNOS TÓZSÉR

In our article we reflect on the claims made by Tamás Ullmann in his 2006 paper *Analytic and Continental Philosophy*. In discussing the differences between contemporary analytic philosophy and phenomenology, Ullmann addresses the fundamental methodological and metaphysical divides between the two approaches as well as their disagreement on the nature of intentionality. He argues that the approach of phenomenology is clearly preferable, and attempts to establish that the ideas of present-day analytic philosophy about intentionality are ridden with preconceptions. Our study examines Ullmann's arguments in depth and shows why his critique is unfounded and his positive theses are highly debatable.

AZ SZTE EGYETEMI KÖNYVTÁR TÁRSADALOMELMÉLETI GYŰJTEMÉNYE  
és az SZTE BTK Filozófia Tanszéke által  
(több hazai és határokon túli partnerintézmény közreműködésével)  
készített és folyamatosan fejlesztett

### MAGYAR NYELVŰ FILOZÓFIAI IRODALOM ELEKTRONIKUS ADATBÁZISA

<http://bodza.bibl.u-szeged.hu:9802/human/philobibl>

1881-től kezdődően rögzíti a magyar nyelven megjelent filozófiai könyvek, illetve időszaki kiadványokban publikált tanulmányok és kisebb-nagyobb közlemények bibliográfiai adatait. Az adatbázis jelenleg mintegy 40 000 (1881 és 2010 között megjelent) magyar nyelvű filozófiai publikáció adatait tartalmazza. Lehetővé teszi a szerző, cím, tárgyszó, 'teljes szöveg' (a keresett kifejezés a bibliográfiai leírás bármelyik elemében előfordulhat, a legnagyobb találati pontossággal itt lehet keresni) és a kiadási adatok stb. szerinti visszakeresést, illetve a hatékony szakirodalmi tájékozódást.

Az adatbázis részét képező *filozófiai folyóirat-repertóriumok, illetve szakirodalmi bibliográfiák* adatállományait szöveges formában külön is elérhetővé teszi.

<http://primus.arts.u-szeged.hu/doktar/filo.html>

A program szakmai irányítója:  
Laczkó Sándor  
[laczkos@bibl.u-szeged.hu](mailto:laczkos@bibl.u-szeged.hu)

Főmunkatárs:  
Balog Éva  
[balogeva@bibl.u-szeged.hu](mailto:balogeva@bibl.u-szeged.hu)



## KÖVETKEZŐ (2010/3) SZÁMUNK TÉMÁJA A KAUZALITÁS

Az oksággal kapcsolatos kérdések az antikvitástól fogva máig a filozófiai elemzések kitüntetett tárgyát képezik, miközben a fogalom számos történeti változáson ment keresztül. A kauzalitás többretegű, gazdag arisztotelészi felfogását, amely a világban található dolgok és események kapcsolatait egy teleologikusan rendezett univerzum részeként vizsgálja, a 17. században a hatáokság kitüntetett szerepe váltja fel. A hagyományos problémákat az okkazi-onalizmus és David Hume óta egy sor újabb kérdés egészíti ki, illetve váltja fel. Mára nyitott kérdés, hogy miben is áll két esemény kauzális viszonya. Elemezhető-e más metafizikai fogalmak segítségével, vagy redukálhatatlan reláció? Egyáltalán, események közötti viszony-e, vagy más relatívumai vannak? Helyes-e Hume intuíciója, amely szerint az okság csupán regularitás; vagy létezik szinguláris kauzalitás is, netán a kauzalitás az emberi cselekvésre épül? Ha a hume-i meglátás helyes, vajon általánosíthatók-e a regularitáselméletek indeterminisztikus folyamatokra, amint azt a valószínűségi kauzális elméletek teszik, vagy modális szituációkra az okság kontrafaktuális elméletei jegyében? Van-e a mentálisnak oksági ereje, és hogyan viszonyul ez a fizikai kauzalitáshoz? Mi az okság viszonya a többi jelentős ismeretelméleti, metafizikai és tudományfilozófiai fogalomhoz: a magyarázathoz, a konfirmációhoz, a természettörvényekhez, a diszpozíciókhoz, a determinizmushoz? Végül, milyen szerepet tölt be a kauzalitás a modern természet- és társadalomtudományokban?

Ezt követő tematikus számaink és a kéziratok leadásának határideje:

*Irracionalitás* (2011/1) január 31.

*Hatalom* (2011/2) május 15.

*Tapasztalat* (2011/3) szeptember 15.

### SZERZŐINK FIGYELMÉBE

A *Magyar Filozófiai Szemle* referált folyóirat, amely csak eredeti kéziratokat közöl, és minden beérkezett tanulmányt két független, anonim bírálóknak ad ki. A kéziratokat elektronikus formában várjuk a [filszemle@webmail.phil-inst.hu](mailto:filszemle@webmail.phil-inst.hu) címre. A teljes példány mellé egy másik, név és önhivatkozások nélküli változatot is kérünk a referensek részére.

A kéziratokat Times New Roman betűtípusban, „rtf” formátumban, körülbelül egyívnyi (40 000 leütés) terjedelemben, csupán a legszükségesebb formázásokkal (például kurziválás) kérjük. A főszövegben a 4-5 soros terjedelmesebb idézetek a törzsszövegtől egy-egy sorral elválasztva, behúva szerepeljenek.

Tanulmányok esetében a hivatkozott szakirodalmat a szövegbe illesztve kérjük feltüntetni: (Kuhn 1999a. 99), vagy lábjegyzetben: Putnam 1975b. 237. Szükség esetén megadható az eredeti megjelenés dátuma és a kötetszám is (Descartes 1644/1996. 8/2:72). Használhatók a széles körben elterjedt rövidítések (AT VIII-2. 72), és klasszikus szövegek esetében a bevett utalási eljárások (Platón, *Phaidrosz* XXIV 245c–d; Mt 9, 13). A rövidítések feloldását kérjük az írások végén, az irodalomjegyzékben megadni.

Az irodalomjegyzéket a következő példának megfelelően kérjük összeállítani:

Önálló kötet:

Locke, John 1690/2003. *Értekezés az emberi értelemről*. Ford. Vassányi Miklós – Csordás Dávid. Budapest, Osiris.

Tanulmány gyűjteményes kötetben:

Henke, Christopher R. – Thomas F. Gieryn 2008. Sites of Scientific Practice: The Enduring Importance of Place. In Edward J. Hackett – Olga Amsterdamska – Michael Lynch – Judy Wajcman (szerk.) *The Handbook of Science and Technology Studies*. Cambridge/MA, The MIT Press. 354–376.

Periodikumban megjelent cikk, tanulmány:

Schnädelbach, Herbert 1993. A filozófia és az embertudományok. Ford. Mezei György Iván. *Magyar Filozófiai Szemle*, 37/2. 1015–1031.

A kéziratokhoz rövid bemutatkozást és angol nyelvű összefoglalót kérünk mellékelni.

# MAGYAR FILOZÓFIAI SZEMLE

2010/2 (54. évfolyam)

A *test* a filozófiának mindig is az egyik alapproblémája volt a test–lélek viszony klaszszikus kérdésétől kezdve a szónak a nem emberi testekre való átviteléig, a mértani testektől a közösség testéig. Gyakran szerepelt a test különböző kozmológiai modellek kulcsszavaként, vagy a politikafilozófia nélkülözhetetlen modelljeként is. A szónak ezeket a használatait természetesen rutinszerűen vádolták azzal, hogy félrevezető metaforák foglyává teszik a (mértani, égi, politikai, állati vagy emberi) testekkel operáló kozmológiai, politikai és egyéb elméleteket, ugyanakkor nemigen tudjuk elképzelni, hogy miképpen lehetne hasonló elméleteket ilyesféle metaforák használata nélkül megalkotni. Természetesen az emberi test az elmével (avagy a lélekkel) való viszonyától és a „test” szó metaforikus használatától eltekintve is a filozófia vizsgáldás tárgyává válhat, gondoljunk akár napjaink testfenomenológiai vizsgálódásaira, akár a bioetikának a filozófián belül is egyre gyakrabban tárgyalt meggondolásaira. A test így a filozófiában egyaránt megjelenik történeti kérdésként és a kortárs diszkusszió problémájaként: aligha képzelhető el a filozófiatörténetnek olyan korszaka, illetve a kortárs diszkusszióknak olyan irányzata, amelynek ne lenne karakteres álláspontja a testről.

ISSN 0025-0090

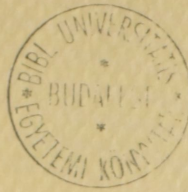


1200 Ft

**GONDOLAT**

**nka**  
Nemzeti Kulturális Alap

P25591 Pn



2011 JAN. 27

A Magyar Tudományos Akadémia  
Filozófiai Bizottságának folyóirata

# MAGYAR FILOZÓFIAI SZEMLE

2010/3 (54. évfolyam)

Okság



## SZERKESZTŐSÉG

Schmal Dániel felelős szerkesztő, Pázmány Péter Katolikus Egyetem  
Ambrus Gergely, Miskolci Egyetem  
Mekis Péter, Eötvös Loránd Tudományegyetem  
Mester Béla, MTA Filozófiai Kutatóintézet  
Ullmann Tamás, Eötvös Loránd Tudományegyetem  
Valastyán Tamás, Debreceni Egyetem

## SZERKESZTŐBIZOTTSÁG

Forrai Gábor, a szerkesztőbizottság elnöke, Miskolci Egyetem  
Almási Miklós, az MTA rendes tagja, Eötvös Loránd Tudományegyetem  
Altrichter Ferenc, University of North Caroline at Wilmington  
Csejtei Dezső, Szegedi Tudományegyetem  
Fehér Márta, Budapesti Műszaki és Gazdaságtudományi Egyetem  
Heller Ágnes, az MTA rendes tagja, Eötvös Loránd Tudományegyetem, The New School for Social Research (New York)  
Hévízi Ottó, MTA Filozófiai Kutatóintézet Lukács Archivum  
Hont István, King's College (Cambridge)  
Kelemen János, az MTA levelező tagja, Eötvös Loránd Tudományegyetem  
Klima Gyula, Fordham University (New York)  
Máté András, Eötvös Loránd Tudományegyetem  
Rózsa Erzsébet, Debreceni Egyetem  
Szabó Zoltán, Yale University  
Tengelyi László, Bergische Universität (Wuppertal)  
Weiss János, Pécsi Tudományegyetem

A szerkesztőség elektronikus postacíme:  
filszemle@webmail.phil-inst.hu  
schmal.daniel@chello.hu

ISSN 0025-0090

Kiadja a Gondolat Kiadó  
Felelős kiadó Bácskai István  
1088 Budapest, Szentkirályi u. 16.  
Telefon: +361-486-1527  
www.gondolatkiado.hu  
http://gondolatkiado.blog.hu

Tördelő: Lipót Éva  
Olvasószerkesztő: Böröczki Tamás  
Lapterv: Andorka Tímea és Pintér László  
Nyomás és kötés: *mondAt Kft.* • Budapest • www.mondat.hu

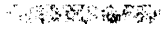
A lap megjelenését a Nemzeti Kulturális Alap  
és a Magyar Tudományos Akadémia Filozófiai Bizottsága támogatja



2011 JAN. 27

# MAGYAR FILOZÓFIAI SZEMLE

2010/3 (54. évfolyam)



A Magyar Tudományos Akadémia  
Filozófiai Bizottságának folyóirata

Okság



# Tartalom

## TANULMÁNYOK

Okság (Schmal Dániel)	5
BETEGH GÁBOR: „A következő ok” Arisztotelész, <i>Metafizika</i> A 3–4	9
BENE LÁSZLÓ: Okság és morális felelősség Plótinosznál	25
SIMON JÓZSEF: Az intelligibilitás okságától az okság intelligibilitásáig A kauzalitás Duns Scotus filozófiájában	45
SZALAI JUDIT: Arisztotelészi-skolasztikus elemek a descartes-i szenvedélytanban	65
E. SZABÓ LÁSZLÓ – GYENIS BALÁZS – GYENIS ZALÁN – RÉDEI MIKLÓS – SZABÓ GÁBOR: Korrelációk kauzális magyarázata	78
HUORANSZKI FERENC: Távolba ható okok	98

## FÓRUM

FARKAS KATALIN: <i>The Subject's Point of View:</i> A könyv összefoglalása	121
MÁRTON MIKLÓS: Vonakodó karteizianizmus?	124
FORRAI GÁBOR: A privilegizált hozzáférés mint a mentális kritériuma: három aggály	133
TŐZSÉR JÁNOS: Ahogy a dolgok tűnnek a szubjektum számára	137
FARKAS KATALIN: Válasz Márton Miklósnak, Forrai Gábornak és Tőzsér Jánosnak	140

## SZEMLE

BOROS GÁBOR: Gábor György – Vajda Mihály (szerk.) <i>A lét hangoltsága. Tanulmányok a tudás sokféleségéről</i>	145
E SZÁMUNK SZERZŐI	149
SUMMARIES	150





## Okság

„Íme egy asztal, s rajta két biliárdgolyó: az egyik áll, a másik gyorsan közeledik feléje.” Talán nincs egyszerűbb és ismerősebb tárgya a filozófiai vizsgálódásoknak, mint az ok–okozati viszony, amely egyszerre alapja a mindennapi életben való tájékozódásnak (miként azt Hume hétköznapi példája sugallja), s a különböző fizikai vagy társadalmi jelenségekkel kapcsolatos tudományos elméletalkotásnak, hiszen – amint ugyancsak Hume állítja – „a *tényeket*” illető következtetéseink mindig az ok és okozat relációján alapulnak, és [...] soha nem tudunk az egyik tárgynak a létezéséből a másikéra következtetni, ha a két tárgy nem áll közvetett vagy közvetlen kapcsolatban egymással” (David Hume: *Értekezés az emberi természetről*. Ford. Bence György. Akadémiai Kiadó. 2006. 438). És mégis, az okság maradéktalanul ismerős jellege dacára máig nyitott kérdés, hogy miben is áll két esemény kauzális viszonya. Elemezhető-e más metafizikai fogalmak segítségével, vagy redukálhatatlan reláció? Melyek e reláció terminusai, és hányan vannak? Tárgyak, események vagy tények közötti viszony-e, esetleg egyéb relátumokkal rendelkezik? Helyes-e Hume intuíciója, amely szerint az okság csupán regularitás; vagy létezik szinguláris kauzalitás is, netán a kauzalitás az emberi cselekvésre épül?

A *Magyar Filozófiai Szemle* 2010/3. számának tanulmányai egy-egy speciális problémán keresztül képet adnak az ok–okozati viszonnal kapcsolatos elméletek történeti alakulásáról a görög filozófia kezdeteitől a kora újkorig, majd az utolsó két tanulmány bepillantást enged a 21. századi vitákba. A kortárs kiindulópont – pozitív előjellel vagy kritikai céllal – legtöbbször a kauzalitás hume-i analízise, így a vizsgálódások legfontosabb viszonyítási keretét ma is az okság egy olyan felfogása alkotja, amely a fizikai világ tudományos leírásának újkori programjához kapcsolódik. E megközelítés több ponton ellentétes azokkal az evidenciákkal, amelyek a korábbi elméletek alapjait képezték egészen a 17–18. századig. Ilyen evidencia a szimultán kauzalitás lehetősége, amelyet Hume határozottan elutasít, vagy az okok Arisztotelész által megjelölt négyféle típusának megkülönböztetése, amelyből a *mainstream* 17. századi elméletek szinte csak a hatóokságnak tulajdonítanak kitüntetett szerepet.

A korábbi, négyes felosztás genezisébe enged bepillantást a tanulmányblokk élén Betegh Gábor „*A következő ok*” című írása, amely az arisztotelészi *Metafizika* A 3–4. részletes elemzésén keresztül azt mutatja meg, hogy Arisztotelész az egyes okfajták megjelenését nem nemzedékről nemzedékre haladó, additív történeti folyamatnak, hanem olyan temptatív lépések sorozatának tekintette, amelyek fokozatosan tették elkülöníthetővé az oksággal kapcsolatos kezdeti intuíció differenciálatlan elemeit. Arisztotelész szerint ugyanis „valamilyen tisztázatlan és kezdetleges formában az okok többféle fajtája is jelen volt már a kezdetektől”. Az elemzés nem csupán Arisztotelész történeti fejtegetéseit rekonstruálja, hanem egyszersmind arra is rámutat, hogy a négy ok – jóllehet metafizikailag megkülönböztethető – egyfajta „minden vagy semmi” kérdésként Arisztotelésznél szorosan összetartozik: nem lehetséges az okság egyes formáiról számot adni oly módon, hogy a többi teljes egészében tisztázatlan marad.

A Hume előtti kauzalitáselméletek egy másik jellegzetes vonása jelenik meg plasztikus formában Bene László írásában (*Okság és morális felelősség Plótinosznál*): a kauzalitás és a szabadság konfliktusa gyakran nem a fizikai folyamatok determinisztikus voltából, hanem az immateriális lélek más, anyagtalan princípiumoktól való függéséből fakad. Bene László az okságprobléma egy ma is centrálisnak számító aspektusát, a kauzalitás és a morális felelősség összeegyeztethetőségének kérdését vizsgálja Plótinosz filozófiájában. Maga a kérdés a hellénisztikus kori vitákban a sztoikus fátumelmélet hatására jelenik meg. A szerző a szakirodalom jelentős részével szemben azt állítja, hogy Plótinosz közelebb áll a sztoikus nézetekhez, mint akár a peripatetikusok indeterminisztikus metafizikájához, akár a középső platonikus szerzők (hagyományos módon interpretált) felfogásához. Plótinosz elfogadja a kozmoszon belüli oksági determináció elvét, s ezzel együtt a szabadság egy olyan kompatibilista elmélete jelenik meg nála, amelyben a döntések a lélek jelleméből, karakteréből vagy diszpozíciójából erednek. A lélek ugyanakkor Plótinosznál nincsen alávetve az anyagi világ determinisztikus folyamatainak, hanem önálló oksági tényezőként tevékenykedik a világban – ami persze nem jelenti azt, hogy cselekvésői ne függnének valamilyen magasabb princípiumtól az emanáció rendjében. Oksági determinizmus és a szabad választáshoz szükséges autonómia a szerző szerint a normatív racionalitás fogalma segítségével kapcsolható össze Plótinosz elméletében.

E Plótinosznál és az újplatonikus szerzőknél meghatározó metafizikai hierarchia átalakulásának lehetünk tanúi Simon József hiánypótló tanulmányát olvasva (*Az intelligibilitás okságától az okság intelligibilitásáig: a kauzalitás Duns Scotus filozófiájában*), amely az oksággal kapcsolatos filozófiai gondolkodás egy kevésbé ismert, ám nagy jelentőségű pillanatát világítja meg. A szerző szerint Duns Scotus elméletében válik el egymástól először jól látható formában az okság önmagában vett elgondolhatóságának kérdése az aktuálisan fennálló kozmosz ontológiai szerkezetére vonatkozó kérdéstől. Amikor Scotus a metafizika tárgyát nem ontoteológiként (vagyis egy okként kitüntetett, tökéletes létező felől) gondolja

el, hanem a másik arisztotelészi megközelítést favorizálva a létezés minimális mozzanatát, a „létezőt mint létezőt” tekinti e tudomány első tárgyának, akkor e fogalommal a lehetséges létezés egy olyan területét írja körül, amelyet kizárólag az ellentmondás-mentesség követelménye határol. Scotus a pusztá lehetőség e háttére előtt keresi az oksági viszonyok megkülönböztető jegyeit. Okról és okozatról szólva annak a lehetőségéről beszélünk, hogy valami valami másból ered (*existens ab alio*). Ám ott, ahol az okságot pusztán formálisan, minden egyébtől elvonatkoztatva tekintjük, a mástól való függés e mozzanata önmagától függés-ként őrzi meg a kauzalitás e lényegi szerkezetét.

A történeti blokk lezárásaként Szalai Judit tanulmánya (*Arisztotelészi-skolasztikus elemek a descartes-i szenvedélytanban*) a kauzális viszony egy speciális, ugyanakkor a descartes-i elméletben közismert nehézségeket okozó esetét, a test–lélek interakció problémáját vizsgálja. Állítása szerint a descartes-i érzékelésemélet e döntő ponton azzal az arisztotelészi eredetű megközelítéssel rokonítható, amelyet Descartes meghaladni kívánt. A pszichofizikai kölcsönhatás ugyanis Descartes szerint egyszerre cselekvés az ok és elszenvedés az okozat részéről, amennyiben tehát észlelési folyamatról van szó, az elmének passzív befogadóként kell viselkednie. Annál különösebb, hogy Descartes az érzékelést gyakran a lélek „odafordulásaként” és a test állapotait „szemlélő” tevékenységeként írja le. A tanulmány tézise szerint a percepció e látszólag aktív aspektusát egy a skolasztikus szerzőktől kritikátlanul átvett mozzanat magyarázza: a „szemlélés” nem valódi aktivitást jelent (vagyis az elme nem rendelkezik olyan intencionális aktussal, amely az agy valamely állapotára irányulna), hanem a formák megragadásának arisztotelianus elméletére utal, s a *speciess*ek befogadásával kapcsolatos skolasztikus nézetek maradványa Descartes-nál.

A tanulmányok sora két kortárs megközelítéssel folytatódik. A *Korrelációk kauzális magyarázata* című tanulmány szerzői arról az idestova tizenöt éve folyó munkájukról adnak áttekintést, amelyet a reichenbachi közös ok elvének nevezett princípium („nincs korreláció kauzalitás nélkül”) tanulmányozása terén folytattak. A közös ok elvének értelmében bármely olyan korreláció, amely nem logikai kapcsolaton alapul, vagy a korrelációban álló események közötti közvetlen kauzális viszonyból, vagy egy harmadik eseményből, a korreláló események közös okából ered. E kutatás során, amely a fogalmak szigorú valószínűségelméleti modellezésére támaszkodott, a szerzők azt vizsgálták meg, hogy a közös ok elvének jelentése, illetve érvényessége miképpen módosul a közös ok fogalmának különféle klasszikus és kvantumos értelmezései és általánosításai mellett, továbbá hogy az elv különféle értelmezései hogyan viszonyulnak a kvantumelmélet úgynevezett rejtett paraméteres kutatásaihoz.

A tanulmányblokk utolsó írása, Huoranszki Ferenc *Távolba ható okok* című munkája a hume-iánus és kontrafaktuális felfogás körül napjainkban zajló viták egyik központi témáját, az okság tranzitivitásának kérdését tárgyalja (ha *A* oka *B*-nek, és *B* oka *C*-nek, akkor *A* oka *C*-nek). A szerző úgy véli, a tranzitivitás

kérdése (a hume-i intuícióktól eltérően) nem vizsgálható a relációba állított partikulárek okságilag releváns tulajdonságaitól függetlenül. Javaslatára szerint e tulajdonságokat a hume-i árus keretet elutasítva célszerű olyan oksági diszpozícióknak tekinteni, amelyek nem manifesztálódhatnak a konkrét fizikai interakcióhoz szükséges feltételek nélkül, s így a mikrofizikai folyamatokat magyarázó oksági erők nem tranzitívak. Ebből ugyanakkor nem arra következtet, hogy a kauzalitás mint olyan nem tranzitív reláció, hanem arra, hogy tranzitív jellege nem magyarázható a mikrofizikai interakciókon keresztül. A tranzitivitás abból az intencionális perspektívából ered, amely a kauzális folyamatok elemeit manipulálható, potenciális eszközöknek tekinti távolabbi célok elérése érdekében.

A szubjektív perspektíva – ha némileg eltérő értelemben is – a *Fórum* rovat írásainak is kulcsmozzanata. A rovat Farkas Katalin *The Subject's Point of View* című 2008-as könyvét mutatja be a kötet tartalmának ismertetésén, három recenzió és a szerző válaszain keresztül. Farkas a kortárs elmefilozófia többségi álláspontjával szemben amellet érvel, hogy az elméről és a személyekről alkotott alapvető meggyőződéseink következetes és kompromisszumok nélküli internalizmust tesznek indokolttá. A kötet tehát azt kívánja megmutatni, hogy az elme megkülönböztető jegyét a privilegizált hozzáférés jelenti, s egy mentális vonás azonosításának az az egyetlen kritériuma, hogy „a világ hogyan jelenik meg az alany számára”.

Lapszámunkat Boros Gábor recenziója zárja az MTA Filozófiai Kutatóintézetének legutóbbi termését bemutató gyűjteményes kötetéről, amely Gábor György és Vajda Mihály szerkesztésében *A lét hangoltsága: Tanulmányok a tudás sokféleségéről* címmel jelent meg idén a Typotex Kiadónál.

*Schmal Dániel*

## „A következő ok”

Arisztotelész, *Metafizika* A 3–4

## I. BEVEZETÉS

Miután a *Metafizika* első két fejezetében Arisztotelész megtette bevezető megjegyzéseit a „bölcesség” (*szophia*) természetéről és tárgyáról, a harmadik fejezet elején rátér elődei okokra és princípiumokra vonatkozó nézeteinek bemutatására. A negyedik fejezet végén azután a következő szavakkal összegzi ennek az áttekintésnek első nagyobb egységét: „Az erre a két okra vonatkozó vizsgálódások tehát az előzőekben, mint mondtuk, eddig jutottak” (985b20–22).<sup>1</sup> Miként a harmadik fejezet elején megtudhattuk, e két ok egyike, az anyagi ok már Thalesz óta, vagy akár még annál is régebben, Homéroszsal kezdve a színen volt. A bölcselkedők viszont csak egy második lépésben kezdtek figyelmet fordítani egy másik okfajta, a mozgató okra, és csak ekkor kezdtek el beszélni egy másik, az anyagi szubsztrátumtól megkülönböztetett princípiumról, amely el tudja látni a mozgató ok szerepét. Tanulmányomban azt szeretném elemezni, hogy miként írja le Arisztotelész e második okfajta megjelenését: milyen megfontolások és kérdések, erők és motívumok vezették elődeit arra, hogy egy második okfajta koncentrálnak, és egy, az előzőektől megkülönböztetett princípiummal álljanak elő.

Mint rövidesen látni fogjuk, Arisztotelész a *Metafizika* első, Alfa könyvének harmadik és negyedik fejezetében többször beszél arról, hogy a második ok és princípium megjelenését milyen kényszerítő erők váltották ki. E megjegyzések alapján az olvasó okkal gondolhatja, hogy Arisztotelész szerint az igazság fokozatos felismerése, és különösképpen a második okfajta felfedezése, determinisztikus. Ezzel szemben azt szeretném kimutatni – részben magára a *Meta-*

E tanulmány a Leuvenben 2008 júliusában megrendezett *Symposium Aristotelicum*on megtartott előadásom lényegesen rövidített és átdolgozott változata. Köszönöm John Cooper, Dorothea Frede, Stephen Menn, Oliver Primavesi, Christof Rapp, Malcolm Schofield, David Sedley, Carlos Steel és Annick Stevens hasznos megjegyzéseit. A magyar változat írásához segítséget kaptam az NKTH kutatói támogatásából (ERC\_HU BETEGH09). A *Metafizika* Alfa könyvének szövegéhez Oliver Primavesi új, megjelenés előtt álló kritikai szövegkiadását használom. A Ross és Jaeger szövegétől való jelentősebb eltéréseket külön jelzem. A görög szövegek fordításait én készítettem.

<sup>1</sup> Ross és Jaeger kiadása szerint ezekben a vizsgálódásokban „elődeink” eddig jutottak.

*fizika* első könyvére, részben más Arisztotelész-szövegekre való hivatkozással –, hogy Arisztotelész szerint az újabb és újabb okfajták felfedezése talán mégsem olyan egyenes vonalú és jól meghatározott folyamat. Miként Arisztotelész maga mondja, a filozófia első lépései botladozóak voltak – ha pedig valaki botladozik, tehetnének hozzá, nem mindig teljesen világos, hogy a következő lépését milyen irányba teszi majd. E korai elméletek Arisztotelész által kimutatott következetlenségei, pontatlanságai és kezdetlegességei azt is jelentik, hogy minden fázis a fejlődés több útját is kínálta. Kérdések és talányok, melyek látszólag az arisztotelészi négy ok egyikére lennének hivatottak vezetni a gondolkodókat, néha, úgy tűnik, egy másik okfajta bevezetéséhez vezetnek, míg egy új típusú princípium tételezése olyan megfontolások alapján is történhet, amelyek akár egy másik princípiumfajtaéhoz is vezethettek volna.

Ehhez kapcsolódik tanulmányom második fő állítása is. Kézikönyvekben gyakran olvashatjuk, hogy Arisztotelész a *Metafizika* első könyvében egy olyan történetet ad elő, amely szerint a legelső bölcsék csak az anyagi okot tételezték, aztán későbbi gondolkodók ehhez hozzáadták a mozgató okot, és így tovább. Ez azonban – miként a következőkben megpróbálom kimutatni – megtévesztően torz képe annak, amit Arisztotelész valójában állít. És ha belegondolunk, valószínűtlen is lenne, hogy Arisztotelész azt tételezze fel, létezik olyan filozófiai elmélet a természetről és a kozmosz létrejöttéről, amelyik pusztán és kizárólag az anyagi okkal – mármint az arisztotelészi értelemben vett, jól meghatározott anyagi okkal – operál. Egy ilyen elmélet képtelen lenne a változásról, de akár az egyes elkülöníthető dolgok létéről is számot adni. A korai preszókratikus elméletekről Arisztotelész természetesen állíthatja, hogy az ezekben az elméletekben szereplő princípiumok leginkább az anyagi ok funkcióit látják el. Viszont ahhoz, hogy ezek az elméletek bármilyen szinten is, de a természet és a természetben folyó változások leírását tudják adni – márpedig a korai ión filozófusok elméletei éppen ezt a célt tűzték ki maguk elé –, szükséges, hogy princípiumaik a mozgató, és bizonyos mértékben a formai ok oksági szerepét is betöltsék. A korai bölcsék princípiumában, az anyagi szubsztrátumban számos okfajta keveredik kezdetleges összevisszaságban. Miként látni fogjuk, így lesz ez a mozgató okkal is: egy elmélet nem tud meglenni pusztán az anyagi és a mozgató okkal, hanem szükségszerűen tartalmazni fogja a cél- és a formai okság kezdetleges elemeit is. Vagyis távolról sem arról van szó, hogy a filozófia korai történetében generációról generációra egymás után jelennek meg teljes vértetben az arisztotelészi okok, hanem sokkal inkább arról, hogy valamilyen tisztázatlan és kezdetleges formában az okok több fajtája is jelen volt már a kezdetektől. A filozófia korai fejlődése Arisztotelész szerint inkább úgy írható le, hogy a gondolkodók egyre tisztábban, de azért még mindig nagyon kezdetleges módon, kezdték megkülönböztetni az oksági funkciókat, és ennek megfelelően újabb és újabb típusú princípiumokat vezettek be, amelyek lehetővé tették az oksági feladatok tisztább elkülönítését.

Harmadik fő állításom pedig az lesz, hogy azok, akik szerint Arisztotelész történeti áttekintésében tendenciózusan és illegitim módon félreértelmezi elődeit a saját elméletének és fogalmi rendszerének rájuk kényszerítésével, alapvetően félreértik a *Metafizika* Alfa könyvének célkitűzéseit.

Ezek az elemzések a *Metafizika* első könyvének pontosabb megértésén túl általánosabb tanulságokkal is szolgálhatnak Arisztotelész okságméleletéről. Azt mutatják ugyanis, hogy a négy okfajta metafizikailag ugyan nem vezethető vissza egymásra, ám episztemológiailag szorosan összefüggnek – mégpedig nem pusztán abban az értelemben, hogy egy teljes oksági leírás mind a négy okot meghatározza,<sup>2</sup> hanem abban az értelemben is, hogy a négy ok elmélete egy „minden vagy semmi” ügy. Egyik okról sem lehet igazán pontos fogalmunk, ha a másik háromról csak ködös elképzeléseink vannak.

## II. A KORÁBBI ELMÉLETEK ELÉGTELEN VOLTÁ

Arisztotelész a második okfajta megjelenéséről szóló részt a következő megjegyzésekkel vezeti be:

Ezen gondolkodók és princípiumok után, mivel ezek elégtelenek voltak a létezők természetének világra hozására, a bölcselkedőket a valóság maga, miként mondtuk, újra arra kényszerítette, hogy kutassanak a következő princípium után. Hisz nem valószínű sem az, hogy akár a tűz, akár a föld, vagy bármi hasonló legyen az oka annak, hogy a dolgok részben szépen és jól vannak, részben pedig ilyen állapotba kerülnek, miként az sem valószínű, hogy bárki is így gondolta volna. Ahogy egyébként az sem lenne helyes, ha valaki ilyen jelentős dolgokat a véletlenre vagy a spontaneitásra bízna (984b8–15).

A bekezdés kettős visszautalással kezdődik: a valóság újra kutatásra kényszerítette a gondolkodókat, és Arisztotelész erről már előzőleg is említést tett. A visszautalás a harmadik fejezet egy előző szövegrészére történik (984a17–29), melyben Arisztotelész arról beszélt, hogy a *pragma* maga, a világ tényei és eseményei, és legfőképpen a változás tapasztalata „mutatták az utat és segédkeztek abban, hogy a gondolkodókat kutatásra kényszerítsék” a változás okáról, „hogy miért is történik a változás és mi az oka.” A *pragma* segítő noszogatása bizonyos tekintetben valóban előrevitte a kutatást, ám a bölcsék válaszaikat továbbra is az anyagi szubsztrátum alapján fogalmazták meg. A változás okára vonatkozó kérdés azonban, legalábbis arisztotelészi nézőpontból, nyilvánvalóan csak egy az anyagi princípiumtól megkülönböztetett princípium megadásával lehet megválaszolni.

<sup>2</sup> Vö. pl. Broadie 2009, különösen 28–29.

ni. Erre azonban várni kellett mindaddig, amíg a valóság maga másodszer is be nem avatkozott,<sup>3</sup> és a gondolkodókat „újra arra kényszerítette, hogy kutassanak a következő princípium után”. A valóságnak és a világ tényeinek tehát kétszer is neki kellett rugaszkodnia, hogy az embereket rávezessék egy a mozgató ok funkcióit ellátni képes princípium bevezetésére. Azt kell hát gondolnunk, hogy a valóság és a tények útmutatása, mi több, kényszerítése, az első körben teljesen eredménytelen és haszontalan lett volna?

Nézzük meg közelebbről, hogy mi is történt az első esetben. Ami magyarázatra szorult, az a változás ténye. Ez vezetett aztán a következő problémához: ha csupán egyetlen fajta princípium, az anyagi szubsztrátum segítségével akarjuk leírni a világot, akkor nem világos, miként tudnánk megmagyarázni a változást, „hiszen a szubsztrátum maga legalábbis nem változtatja meg önmagát” (984b21). A korai ión monisták és az eleata filozófusok különféleképpen reagáltak a valóság felől érkező nyomásra. Az eleaták, úgy tűnik, megértették a probléma jelentőségét, viszont választ nem találva inkább bedobták a törülközőt, és feláldozták a változást, „kudarcot vallva a második ok után történő kutatásban”. A változás tapasztalata volt hivatott rávezetni őket egy a mozgás okaként funkcionáló princípium bevezetésére, ám ők, nem lelvén a választ, inkább lemondtak az empirikus tapasztalat érvényességéről, és tagadták a változás tényét. Ez azonban világos, hogy zsákutca.

Úgy tűnik, az ión monisták ezzel szemben nem is ébredtek rá a kérdés jelentőségére, és „a legkevésbé sem zavartatták magukat” (284a29). Nem tagadták a változás tényét, de nem is különösebben érdekelte őket a változás forrása és oka. Pontosabban úgy vélték, épp elég, ha feltételezik, hogy az anyagi szubsztrátum önmagában is dinamikus, és így a változás magyarázatához nincs szükség semmilyen, az anyagi szubsztrátumtól független entitás tételezésére. A kérdést megoldottnak tekintették azzal, hogy az anyagi szubsztrátumon belül tettek bizonyos megkülönböztetéseket, hogy így aztán az ellentétes minőségek egyikéről, mondjuk a melegről, vagy az elemek egyikéről, mondjuk a tűzről mondják azt, hogy ezek mozgatják ellentétüket, illetve a többi elemet.

A *Fizika* első könyvének negyedik és ötödik fejezete érdekes részletekkel egészíti ki az itt kirajzolódó képet. Arisztotelész ugyanis ott emellett érvel, hogy az összes olyan korai gondolkodó, aki magyarázni akarta a változást, az ellentétekre építette elméletét. Ez pedig nem pusztán azokra igaz, akik egynél több anyagi princípiumot tételeztek, hanem a monistákra is, akik aztán az egyetlen szubsztrátum sűrűsödésével és ritkulásával magyarázták a létezők sokaságát. Mi több, még maga Parmenidész is hajlandó volt elfogadni az ellentétek létét, már

<sup>3</sup> A bekezdésben az *alétheia* szót fordítottam „valóság”-nak. Ezt egyrészt alátámasztja az, hogy nem az igazság, hanem a valóság az, ami további kutatásra sarkallja a kutatókat, másrészt és még inkább az, hogy az *alétheia* a *pragma* szinonimájaként szerepel. Arról, hogy az *alétheia* Parmenidésznél is „valóság”-nak (*reality*) fordítandó, lásd Palmer 2009. 190–193.



amennyiben hajlandó volt elfogadni a változás tényét (azaz a *doxát*), és a tüzet és a földet, azaz a meleg és a hideg ellentétpárját tekintette princípiumoknak (*Fizika* I 5, 188a19–27). Arisztotelész ezek után azzal folytatja, hogy ennyiben teljesen igazuk is volt a korai gondolkodóknak, hiszen az ellentétek ténylegesen alapvető szerepet játszanak a változásban, illetve annak megértésében, hogy a változások nem összevissza, hanem jól rendezett módon, ellentétből ellentétbe történnek. Ám Arisztotelész még egyszer visszatér az ellentétek szerepével kapcsolatos *communis opinio*hoz:

Miként mondtam, eddig a pontig a gondolkodók többsége gyakorlatilag egyetért velünk, hiszen mindannyian az ellentétekkel azonosították az elemeket és azt, amit princípiumnak neveznek, még ha nem is okolták meg, hogy ezt miért teszik, hanem *mintha maga a valóság kényszerítette volna őket erre* (*Fizika* I 5, 188b27–30).

A megfogalmazás egyértelműen egyezik a *Metafizikában* olvasottakkal. Az ellentétek fontosságát és a kép koherenciáját tovább erősíti, amit a *Metafizika* első könyvének egy későbbi fejezetében olvashatunk: Parmenidész arra kényszerült, hogy kövesse a jelenségeket, és feltételezze két ellentétes princípium, a tűz és a föld létét (*Met.* A 5, 986b27–987a2).

Mindezen szöveghelyek tehát azt mutatják, hogy a valóság és a tények tapasztalata már az első esetben sem volt teljesen hatástalan: arra szorította rá a filozófusokat, hogy elméletükben helyet találjanak az ellentéteknek, még akkor is, ha egyébként monista intuícióik voltak. Figyelemre méltó azonban az, hogy Arisztotelész a *Fizika*, illetve a *Metafizika* első könyvében más-más módon bontja ki a történetet. A *Metafizikában* a meleg és a hideg, illetve a tűz és az egyéb elemek közötti ellentét annak a felismerésnek az előhírnökeként szerepel, hogy a változás okaként egy független princípiumot kell tételeznünk. Még ha e korai böleselkedők nem is jutottak el odáig, hogy egy valóban különálló princípiummal magyarázzák a változást, az ellentétek, illetve az elemek egyikét kvázi-mozgatókként kezelték: „úgy használják a tüzet, mint aminek olyan a természete, hogy képes mozgatni” (*Met.* A 3, 984b6–7).

Ezzel szemben a *Fizikában* Arisztotelész azt magyarázza, hogy ezek a korai gondolkodók ugyan nem végezték el a szükséges fogalmi tisztázásokat – nem adtak számot arról, hogy az ellentétek pontosan miért is olyan fontosak, s hogy mi is a viszonyuk a szubsztrátumhoz, ami persze azt is jelenti, hogy a szubsztrátumról magáról sem volt világos fogalmuk –, mindezek ellenére valami alapvetőre mégiscsak rájöttek, amikor az ellentéteket tették meg princípiumoknak vagy a princípium elsődleges módosulásának. Ez egy felettébb ígéretes belátás, hiszen az ellentéteknek valóban alapvető szerepük van a változás magyarázatában. Azonban – és ez a lényeges számunkra – a *Fizikából* az derül ki, hogy ha ezt a belátást módszeresen továbbgondoljuk, és megteesszük a szükséges fogalmi

tisztázásokat, akkor nem az anyagi ok és a mozgató ok közötti megkülönböztetéshez, hanem az anyag és a forma (és a priváció) közötti megkülönböztetéshez jutunk.

A *Metafizika* első könyvében szereplő megfogalmazások alapján azt gondolhatnánk, hogy mihelyt a gondolkodókban felsejlett az első ok, az anyagi szubsztrátum gondolata, a változás tapasztalata és az ezzel kapcsolatban felmerülő kérdések és talányok szükségszerűen vezettek a mozgató ok bevezetéséhez. A *Fizika* párhuzamos szöveghelye azonban világossá teszi, hogy legalább két tényező volt jelen már a kezdetektől: nem pusztán a mozgató ok problémája, hanem általánosabban az ellentétek szerepe a változás leírásában, és ily módon a formai ok távoli gondolata is. E korai bölcsék csupán egyetlen fajta princípiummal dolgoztak, ám a valóság és mindenekelőtt a változás tapasztalata abba az irányba vezette őket, hogy megkülönböztetéseket tegyenek az anyagi szubsztrátumban. Ezeket a megkülönböztetéseket aztán a változások számos – és valójában túllontúl sok – aspektusának magyarázatához használták fel; ezzel kívántak számot adni a változások rendezett voltáról, forrásáról, és így tovább. Mivel ezen a ponton mindezen okságilag releváns aspektusok egymáshoz és a szubsztrátumhoz való viszonya nem volt tisztázva, ez az állapot több okfajta későbbi megjelenését is megelőlegezte. Ha pedig ez így van, akkor az első epizód végén még nem is lehetett olyan világos, hogy melyik is lesz az a bizonyos „következő ok”.

A történet első részének végén tehát ott tartunk, hogy egyesek tökéletesen meg voltak elégedve részeredményeikkel, és „a legkevésbé sem zavartatták magukat”, míg mások, az eleaták, bedobták a törülközőt. Az Arisztotelész által használt képet alkalmazva úgy is fogalmazhatunk, hogy a „valóságnak” újra elő kellett állnia valamivel, hogy elmozdulás történjen erről a holtpontról. És tényleg, a valóság tapasztalatában valami új állt elő: a gondolkodók ettől a ponttól kezdve már nem pusztán a mozgásra és a változásra koncentráltak, hanem érdeklődésük előterébe most már a folyamatok és állapotok inherensen jó és szép volta került. Miként Arisztotelész fogalmaz: „Hisz nem valószínű sem az, hogy akár a tűz, akár a föld, vagy bármi hasonló legyen az oka annak, hogy a dolgok részben *szépen és jól vannak, részben pedig ilyen állapotba kerülnek*, miként az sem valószínű, hogy bárki is így gondolta volna” (984b11–14). Fogadjuk el, legalább feltételesen, hogy a tűz vagy valami hasonló képes mozgatni a dolgokat, és akár még azt is, hogy az ellentétes fizikai minőségek garantálják, hogy a változások rendezetten zajlanak – mindezen megkülönböztetések azonban még mindig nem lesznek képesek magyarázni a dolgok jó, szép és kifinomult voltát. Ami új elem a valóság tapasztalatában, az éppen ez: a dolgok nem pusztán mozognak és változnak, de célszerűek, jók és szépek, illetve ilyenné válnak. Úgy tűnik, pusztán a mozgás és a változás tapasztalata önmagában nem volt elegendő ahhoz, hogy a gondolkodók egy önálló princípiumot vezessenek be a megfelelő oksági szerep betöltésére, azaz a mozgás eredetének magyarázatára. Hogy ez megtörténjen, a valóságnak jónak és szépnek kellett feltűnnie a gondolkodók

szemében. Ám ez a mozzanat már önmagában is jelzi, hogy azok, akiknek kutatásait elsődlegesen az mozgatta, hogy kimutassák a jó és a szép forrását és okát, nem pusztán és nem is elsősorban a mozgás forrására és eredetére – azaz a mozgató okra – voltak kíváncsiak.

### III. A KÖVETKEZŐ PRINCÍPIUM

Amikor tehát valaki azt mondta, hogy az ész, csakúgy, mint az állatokban, jelen van a természetben is a világregd és a mindenség elrendezésének okaként, akkor ez az ember józanságával tűnt ki elődei ötletszerű megnyilatkozásaihoz képest. Annyit tudhatunk, hogy Anaxagorasznak gondja volt arra, hogy explicit módon fogalmazzon ezekről a dolgokról, de a klazomenai Hermotimoszról is azt állítják, hogy már korábban is ilyen értelemben beszélt. Azok tehát, akik így nyúltak a kérdéshez, azt állították, hogy a dolgok jó állapotának oka a létezők princípiuma, és ugyanakkor az a fajta ok, amelyből a dolgok mozgása ered (984b15–22).

Miként a bekezdés első mondatából kiderül, Arisztotelész szerint az új eredmények két egymással összefüggő tényezőn alapultak. Egyrészt azon, hogy a kutatás középpontjába az egyedi létezők és jelenségek magyarázata helyett a világegész rendjének kérdése került. Másrészt pedig azon, hogy a kozmoszban mutatkozó rend és célszerű működések magyarázatához az élőlények célszerű és észszerű viselkedését vették alapul.

Az persze távolról sem nyilvánvaló, hogy a bennünket körülvevő dolgok és jelenségek olyan nagyon jó állapotban lennének, szépen működnének, és mindig, vagy legalábbis többnyire jó irányba változnának. Ellenben ha van egy olyan területe a világnak, ahol a dolgok mégiscsak kiszámítható módon, rendben mennek, és aminek a szépsége nem szűnik csodálattal eltölteni a szemlélőt – nos, akkor ez az égi szféra. Ha tehát valaki meg is próbálta a különböző összetett létezőket és jelenségeket, sőt még akár az élőlények felépítését is pusztán az anyagi összetevők kölesönviszonyaival magyarázni, és így „a véletlenre vagy a spontaneitásra bízni”, egy ilyen magyarázati módot szinte lehetetlennek tűnik a szépen, ésszerűen és harmonikusan működő világmindenségre kiterjeszteni. Miként Arisztotelész fogalmaz az *Állatok részei* című traktátusában, „a rendezettség és a jól meghatározottság ugyanis lényegesen nyilvánvalóbb az égbolt esetében, mint a mi környezetünkben, míg a halandók között az látszik inkább, hogy a dolgok hol így, hol úgy, véletlenszerűen történnek” (641b18–19). A dolgok jó és szép voltának magyarázata akkor válik igazán égető kérdéssé, amikor tekintetünket a csillagos égbolt felé fordítjuk. Arisztotelész a *Meteorológia* első könyvének első fejezetében egyébként ki is mondja, hogy a természet részletes vizsgálatának az égi szféra vizsgálatával kell kezdődnie, hogy csak a következő

lépésben térjünk rá a halandó lényekre.<sup>4</sup> Az égbolt vizsgálata tehát módszertani elsőséget élvez. Figyelemre méltó, hogy bár Arisztotelész a *Metafizika* első könyvének második fejezetében azt mondta, hogy az égbolt folyamatosan a rácsodálkozás tárgya, és így a filozófia forrása, a kozmikus perspektíva, az égbolt rendezettségének magyarázata valójában csak most, az ész mint princípium bevezetésével foglalkozó bekezdésben jelenik meg. Anaxagorasz (és talán Hermotimosz)<sup>5</sup> „józansága” tehát nem pusztán, illetve talán nem is elsődlegesen tanai tartalmára, hanem inkább az általa választott módszerre és perspektívára vonatkozik – nem véletlen, hogy éppen ez vezetett ehhez a jelentős előrelépéshez a bölcsesség történetében.<sup>6</sup>

Ám még ha az égbolt szépsége és rendje messze felülmúl is mindent, amit saját környezetünkben tapasztalhatunk, amikor számot kívánunk adni e felsőbb rendezettség és szépség okáról, a legjobb, amit tehetünk, még mindig az, hogy azokra a jelenségekre hivatkozunk, amelyek a Hold alatti világban a legnagyobb rendezettséget és célszerűséget mutatják – ez pedig az értelmes élőlények célirányos viselkedése. Ezen a ponton hivatkozhatunk egy, már az előző epizódban felmerült, de akkor elsikkadt kérdésre: „hiszen a szubsztrátum maga legalábbis nem változtatja meg önmagát, például sem a fa, sem a bronz nem okozza egymás változását, ahogy nem is a fa készíti az ágyat, és nem is a bronz a szobrot, hanem valami más a változás oka” (984a21–25). Nem a fa hozza létre az ágyat, és nem is a bronz a szobrot, hanem az értelmes emberi lények szaktudása. Az, hogy a kozmikus rendet is az emberi értelmes cselekvések és szaktudások analógiájára kell magyaráznunk, hasonlóképpen hangsúlyos az *Állatok részei* imént idézett első fejezetében is:

Hiszen világos, hogy miként a mesterségesen létrehozott dolgokban ott a mesterség, ugyanígy magukban a világ dolgaiban (*pragma*) is van egy másik hasonló princípium és ok, amelyet – miként a meleget és a hideget is – a mindenségből nyerünk. Éppen ezért sokkal ésszerűbb feltételezés, hogy az égbolt maga egy ilyen ok hatására jött létre – már ha valóban egyszer létrejött –, és egy ilyen ok révén létezik, mint hogy a halandó élőlényeket egy ilyen ok hozta létre (641b12–18).

<sup>4</sup> *Meteorológia* I 1, 339a5–9. A szövegrész részletes elemzéséhez, különös tekintettel arra a kérdésre, hogy az égbolt vizsgálata miért elsődleges, lásd Burnyeat 2004. 13–24, Falcon 2005. 2–7. és Sedley 2007. 196–197.

<sup>5</sup> Hermotimoszról a tanulmány hosszabb, angol nyelvű változatában részletesen beszélek.

<sup>6</sup> Ezen a ponton érdemes megjegyeznünk, hogy míg a modern filozófiatörténet-írás egyértelműen Parmenidésztekint a preszókratikus filozófia belső korszakhatárának, addig mind Platón a *Phaidónban*, mind Arisztotelész (nyilvánvalóan a *Phaidóntól* nem függetlenül) a *Metafizika* első könyvében Anaxagorasz tulajdonítja ezt a szerepet.

Az érv nagyon sűrített, és a görög szöveg (miként a fordítás is) igencsak nehézkes. Azonban a részleteket jól kiegészíthetjük azon szövegek alapján, amelyek minden valószínűség szerint az arisztotelészi megfogalmazás háttérében állnak: Xenophón: *Emlékeim Szókratészről* I. 4. 1 és Platon: *Philebosz* 28c–30d.<sup>7</sup> A kulcsmozzanat az a megjegyzés, hogy ezt a bizonyos okot ugyanúgy nyerjük a mindenségből, mint ahogy az alapvető fizikai minőségeket és elemeket.<sup>8</sup> Mind Xenophónnál, mind a *Philebosz*ban ez az ok az ész (*nusz*), és ez az ok az, amely a kozmikus szinten a mindenség rendjéért, a mozgások rendezettségéért felelős. A párhuzamot csak tovább erősíti, hogy Arisztotelész az *Állatok részei* első fejezetének előző bekezdésében éppen azt a kérdést vizsgálta, hogy vajon tud-e mozgatni az ész. És ugyanez természetesen a *Metafizika* első könyvének általunk vizsgált szövegrészével is csak még szorosabbá teszi a kapcsolatot, hiszen itt Arisztotelész éppen azt tárgyalja, hogy az ész mennyiben lehet a világrendben mutatkozó szépség és jószág, illetve általában véve a mozgás oka.

Az *Állatok részeiből* most idézett bekezdés azonban egy olyan nagyobb érvmenet része, melynek egyes részletei ugyan vita tárgyát képezik, ám amelyek végpontját egyértelműen a természetben mutatkozó cél-okság jelenti. Pontosabban, a gondolatmenet konklúziójában Arisztotelész két okfajta különböztet meg, melyek közül az egyik a cél-ok, míg a másik a „szükségyszerűségből eredő ok” (ami leginkább az anyagi oknak feleltethető meg). Az *Állatok részeinek* első fejezete tehát azt teszi nyilvánvalóvá, hogy a kozmosz és a csillagos ég rendezettségének és szépségének a tapasztalata, illetve ennek a tapasztalatnak az ész kozmikus szerepével való magyarázata elvezethet az anyagi ok és a cél-ok megkülönböztetéséhez<sup>9</sup> – azaz nem feltétlenül az anyagi ok és a mozgató ok közötti viszony tisztázásához járul hozzá, mint ahogy azt a *Metafizika* Alfa könyve alapján gondolhatnánk. Úgy is fogalmazhatnánk, hogy az *Állatok részeinek* első fejezete úgy viszonyul a „valóság” második próbálkozásához, ahogy a *Fizika* első könyve viszonyul a „valóság” első segítő közbélépését leíró részhez: azt mutatja, hogy a valóság tapasztalata által a gondolkodókra kényszerített megfontolások nem egyértelműen mutatnak egy bizonyos újabb okfajta, illetve az ahhoz tartozó princípium bevezetése felé.

Mindezek alapján már világosabban értjük az elemzésünk tárgyát képező bekezdés utolsó mondatát is: „Azok tehát, akik így nyúltak a kérdéshez, azt állí-

<sup>7</sup> Arról, hogy az *Emlékeim Szókratészről* fontos vonatkoztatási pont Arisztotelész számára saját természetes teleológiájának kidolgozásában, lásd Johnson 2005. 115–117. A xenophóni és a *Philebosz*ban található érv viszonyáról lásd Frede 1997. 215 és Dorion 2000.

<sup>8</sup> Megjegyzendő, hogy a szövegrészben szereplő meleg és hideg nem pusztán minőségek, hanem éppen azok a kvázi-princípiumként szereplő ellentétek, amelyekről a *Fizikában* olvashatunk.

<sup>9</sup> Az a tény, hogy a kozmikus észről szóló érv a cél-ok természetben betöltött szerepére vonatkozó konklúzióhoz vezet, fontos pontokon kapcsolódik az első, mozdulatlan mozgatónak a *Metafizika* Lambda könyvében kidolgozott tanához. Ennek a párhuzamnak az elemzése azonban túlzó e tanulmány keretein.

tották, hogy a dolgok jó állapotának oka a létezők princípiuma, és *ugyanakkor* az a fajta ok, amelyből a dolgok mozgása ered.” Amit ezek a gondolkodók, Anaxagorasz és, mint majd látni fogjuk, Empedoklész nem tettek meg, az éppen annak tisztázása, hogy ez a két oksági szerep – a jó okának lenni és a mozgás okának lenni – miképp viszonyul egymáshoz.

Mielőtt azonban rátérnénk Empedoklészre, érdemes felfigyelnünk arra is, hogy a mozgató ok és a cél-ok viszonyának kérdése valójában már az egész történeti áttekintés bevezetésében is megjelent. Amikor Arisztotelész a harmadik fejezet elején visszaül a négy ok *Fizikában* bevezetett elméletére, akkor az anyagi és a formai ok említése után így folytatja: „[H]armadjára a változás forrása [vagyis a mozgató ok – B. G.], és negyedszerre *az ezzel ellentétes ok*, az »ami végett« és a jó [vagyis a cél-ok – B. G.], hiszen ez minden létrejövés és változás végpontja.” Az okok és a princípiumok kutatása második epizódjának főszereplői, Anaxagorasz és Empedoklész éppen ezt a viszonyt nem tisztázták, hanem a változás forrását azonosították a jó okával, illetve magával a jóval. A jó volt vizsgálódásuk fókuszában, ám nem értették meg, hogy a jó nem a változás kiindulópontja, hanem éppen a végpontja és a célja, és így éppen az ellentéte a mozgató oknak.

#### IV. A JÓ ÉS A ROSSZ DOLGOK OKA

Arisztotelész a második princípiumtípus tárgyalását egy rövid kitérével folytatja (984b22–31). Ebben azt vizsgálja, hogy Hésziodosz és Parmenidész, vagyis azok a szerzők, akik Erószt, a szerelmi vágyat teszik meg princípiumnak, vajon nem tekinthetők-e a mozgató ok történetének szereplőiként. Ennek a kérdésnek – amelyet Arisztotelész végül megválaszolatlanul hagy – az elemzésére most nem keríthetek sort.<sup>10</sup> Vegyük fel hát a szöveg fonalát ott, ahol Arisztotelész rátér arra a szerzőre, Empedoklészre, aki az Anaxagoraszéhoz mérhető előrelépést tett az újabb princípiumfajta kidolgozásában.

Ám mivel azt is észrevették, hogy a jó dolgok ellentétei is megvannak a természetben – nemcsak a rend és a kiváló, hanem a rendezetlen és a gyalázatos is, sőt, hogy a rossz dolgok számosabbak, mint a jók, és a hitványak, mint a szépek –, valaki más bevezette a barátságot és a viszályt, egy-egy okot a dolgok két fajtájának. Mert ha végiggondoljuk, hogy Empedoklész valójában mit is akart mondani, és nem arra figyelünk, hogy úgy fejezi ki magát, mint egy gyerek, aki még nem tudja rendesen megformálni a sza-

<sup>10</sup> A tanulmány hosszabb angol nyelvű változatában részletesen tárgyalom ezt a szövegrészt is.

vakat,<sup>11</sup> akkor azt találjuk, hogy a barátság a jó dolgok oka, míg a viszály a rosszaké. Így tehát feltehetőleg helyesen fogalmaznánk, ha azt mondanánk, hogy Empedoklész bizonyos tekintetben kimondta, illetve, hogy ő volt az első, aki kimondta, hogy a jó és a rossz princípiumok, már amennyiben a jó maga minden jó dolgok oka, [a rossz pedig a rossz dolgok oka]<sup>12</sup> (984b31–985a10).

Anaxagoraszt az vezette korszakos felfedezésére, hogy arra koncentrált, ami a legnyilvánvalóbban rendezett és szép a világban – ám eközben mintha megfedelkedett volna a valóság sötétebb oldaláról. Empedoklész sokkal jobban – sőt feltehetőleg túlon túl erősen – érzékelte a világ ezen aspektusát, és ezért a rossz dolgok okát, a rosszat magát is megtette önálló princípiumnak. Ha most egy pillanatra eltekintünk attól, hogy a negativitás princípiummá emelése valóban jó gondolat-e – a *Metafizika* Lambda könyvének tizedik fejezetében Empedoklész megkapja a magáét, amiért a rosszat princípiumnak tekinti<sup>13</sup> –, nagyon úgy tűnik, hogy újítása egy az okság természetével kapcsolatos alapvető megérzésen alapul.

Empedoklész ugyanis arra jött rá, hogy ellentétes okozatoknak ellentétes okai kell hogy legyenek. Arisztotelész persze azt is hozzáteszi, hogy a belátást, amely szerint rendszerszerű összefüggés van a két típusú okozat és a két okként szereplő princípium között, Empedoklész nem fogalmazza meg, és nem is alkalmazza egyértelműen. Például a viszály, azaz a rossz princípium hozza létre a kozmosz alapstruktúráját, tehát valami jó és rendezett dolgot, míg a barátság egységesítő működése egybeolvaszt, és ily módon megsemmisít számos olyan dolgot, amit Arisztotelész inherensen jónak tekint. Ha azonban eltekintünk ezektől a következetlenségektől, s türelmesen végighallgatjuk és végiggondoljuk, amit Empedoklész mond, akkor világossá válik, hogy mire is kellett gondolnia. Az a gondolat pedig, hogy a jó dolgokat egy jó dolog, a rossz dolgokat pedig egy rossz dolog okozza, előfutára lehet az okság egy Platón, Arisztotelész és az antikvitás későbbi filozófusai által széles körben elfogadott és alkalmazott alapelvének. Ez az alapelv, amelyet a kortárs szakirodalom az „okági szinonímia elvének” nevez, azt mondja ki, hogy az oknak magának rendelkeznie kell azzal a tulajdon-

<sup>11</sup> A szöveg különböző nyelvű fordításaiban a *pszellidzomai* igét általában a „dadog”, illetve „selypít” megfelelőivel szokás visszaadni. Itt azonban nem felnőttkori beszédhibáról van szó, hanem arról, ahogy egy gyermek még nem képes jól megformálni az egyes fonémákat – ezért használok fordításomban bővebb körülírást. Ugyanezt az igét használja Arisztotelész elődei nem tiszta, megformálatlan beszédmódjára az Alfa könyv utolsó fejezetének összefoglalásában. Vö. még *Historia animalium* 536b8 és Platón: *Gorgiasz* 485b–c.

<sup>12</sup> A szögletes zárójelben szereplő szavak szerepelnek az Alfa kéziratcsalád legtöbb kódexében, azonban a kiadók túlnyomó többsége, így Bonitz, Christ, Ross és Jaeger is, későbbi értelmező betoldásnak véli.

<sup>13</sup> Lásd különösen 1075b7–8.

sággal, amelyet létrehoz az okozatban.<sup>14</sup> Empedoklész csak a jó és a rossz dimenziójára koncentrált, ám az elv, amelyet erre az ellentétpárra alkalmazott, könnyen általánosítható. Amint pedig az elvet kiterjesztettük, a mozgató (vagy ható) okság egy Arisztotelész szerint is központi fontosságú alapelvét értettük meg. Természetesen ahhoz, hogy ebben az irányban megtehessük a következő lépéseket, az arisztotelészi értelemben vett formával és annak a változás szubsztantumához való kapcsolatával kellene tisztába kerülnünk. Éppen ott kellene tehát felvennünk a kutatás fonalát, ahol a *Fizika* első könyve szerint a változásokat az ellentétes minőségekkel magyarázó korai filozófusok abbahagyták a vizsgálódást. (Mindezt Arisztotelész persze nem fogalmazza meg ilyen egyértelműen, ám ha türelmesen végigolvassuk és végiggondoljuk különböző szövegeit, akkor világossá válik, hogy erre kellett gondolnia.)

## V. ÖSSZEFOGLALÁS ÉS ÉRTÉKELÉS

A négy elemfajta elméletének bevezetésével Arisztotelész számára Empedoklész jelentette az anyagi ok előtörténetének csúcspontját. Úgy tűnik, a második okfajta kutatásában is ő jutott a legtovább a két ellentétes, de egyaránt mozgató okként funkcionáló princípium tételezésével. Így Arisztotelész már rá is térhet a természetfilozófusok ezen első nagy csoportjának általános értékelésére:

Úgy látszik hát, hogy ezek a gondolkodók az előzőekben mondtak szerint és eddig a pontig jutottak el a természetről szóló értekezésünkben megkülönböztetett okok közül kettővel kapcsolatban, úgy értem, az anyaggal és a mozgás eredetével kapcsolatban – ám mindezt zavarosan és egyáltalán nem világosan tették, hasonlóan a gyakorlatlan harcosokhoz. Hiszen azok is gyakran szép ütésekkel visznek be, ahogy ellenfelük körül táncolnak, de mindezt valódi ismeret nélkül teszik, miként ezekről is az látszik, hogy nem is tudják, mit beszélnek. Mert hát nyilvánvaló, hogy szinte semmire sem használják ezeket az okokat, vagy talán csak ha egy kicsit (985a12–18).

Arisztotelész ebben a bekezdésben összegzi az áttekintés eddigi eredményeit és mondja ki explicit módon, hogy az a két ok, amellyel a gondolkodók ezen csoportja foglalkozott, az anyagi és a mozgató ok. A „második okot” persze már a valóság első erőfeszítését leíró részben is a mozgató okkal azonosította. Ám miként láthattuk, az egész második epizód, melynek Anaxagoras és Empedoklész voltak a főszereplői, legalább ennyire szólt a jó okáról és a jóról magáról. Ez vált különösen nyilvánvalóvá, amikor Arisztotelész kimondta, hogy az empedoklészi

<sup>14</sup> Lásd pl. Bodnár–Pellegrin 2006. 279–281.



barátságról állíthatjuk azt, hogy az a jó maga. Tekintetbe véve, hogy Arisztotelész a harmadik fejezet elején a cél-okot úgy határozta meg, hogy „az »ami végett« és a jó” (983a31–2), joggal gondolhattuk volna azt is, hogy ezek a filozófusok, ha homályosan is, de valameddig eljutottak a cél-ok felfedezésében is. Azzal kapcsolatban azonban, hogy Anaxagorasz és Empedoklész valamilyen mértékben, illetve valamilyen tekintetben a cél-okkal is foglalkozott volna, semmilyen megjegyzést sem hallunk az Alfa könyv 3–4. fejezetében. Arisztotelész először majd csak a hetedik fejezetben, a történeti áttekintés összegzésében fogja elismerni, hogy ha mégoly korlátozott mértékben is, de az ész és a barátság – most már mindkettő azonosítva a jóval magával – valóban releváns a cél-ok kérdésében is. Mi több, itt tudjuk meg azt is, hogy az Arisztotelész előtti összes filozófus összes princípiuma közül a platóni Egy mellett valójában csak ez a két princípium érdemel említést a cél-ok előtörténetében. Arisztotelész itt mondja ki azt is, hogy Anaxagorasz és Empedoklész princípiumai azért nem tudnak cél-okként funkcionálni, mert még ha azonosíthatók is a jóval magával, semmilyen értelemben sem tekinthetők a változások céljának, annak, „ami végett” a változás végbe megy. Egyrészt világos, hogy Arisztotelész magának akarja fenntartani a cél-ok szabadalmát. Másrészt, a hetedik fejezetben Anaxagoraszról és Empedoklészről tett *utólagos* megjegyzések nélkül értelmezhetetlen maradna a könyvet lezáró tizedik fejezet összegzése, amely szerint az elődök valamilyen értelemben mind a négy okról, azaz a cél-okról is beszéltek (993a14–15).

Az a mód, ahogyan Arisztotelész a negyedik fejezetben végül mégiscsak egyértelműen a mozgató okhoz rendeli Anaxagorasz és Empedoklész princípiumait, véleményem szerint jól összevehető a szerencsének és a spontaneitásnak a *Fizika* második könyvében szereplő tárgyalásával. E hosszas tárgyalás zárómondata szerint „mindkettő a mozgás forrása oktípushoz tartozik” (198a2–3). Ám bármily határozottnak és egyértelműnek tűnik is ez a megfogalmazás, az már egyáltalán nem ilyen bizonyos, hogy a megelőző tárgyalás felhatalmazza Arisztotelészt egy ilyen magabiztos konklúzióra. És valóban, egyes kommentátorok mellett érvelnek, hogy bármit mondjon is Arisztotelész ebben az összegző mondatban, elemzése azt mutatja ki, hogy a véletlen és a spontaneitás nem a mozgató, hanem a cél-ok körébe tartozik.<sup>15</sup> Vagyis egyáltalán nem olyan egyértelmű, hogy a véletlen és a spontaneitás melyik oktípus alá tartozik, ahogy – miként láttuk – egyáltalán nem olyan egyértelmű, hogy az anaxagorasz-i ész és az empedoklészi barátság (és bizonyos értelemben a viszály) végeredményben a mozgató ok vagy a cél-ok alá sorolandó.

Arisztotelész számára azonban az elsődlegesen fontos mindkét esetben az, hogy ezek a jelenségek, illetve oksági hatással rendelkező princípiumok ide vagy oda, de valahová besorolhatóak legyenek. Hiszen miután kijelentette a véletlenről és a spontaneitásról, hogy „mindkettő a mozgás forrása oktípushoz tartozik”,

<sup>15</sup> Lásd pl. Judson 1991. 79–80.

csak ezt követően mondhatja ki, hogy „[n]yilvánvaló tehát, hogy vannak okok, és hogy a számuk annyi, amennyit megállapítottunk” – azaz sem nem több, sem nem kevesebb, mint négy. Hiszen Arisztotelész éppen azért kezdett bele a véletlen és a spontaneitás tárgyalásába, hogy vizsgálat alá vegye azt a széles körben elterjedt – és mint végül kiderül, alapvetően helytálló – nézetet, mely szerint a véletlen és a spontaneitás oksági funkcióval rendelkező tényezői a világnak. Így csak akkor tudja kijelenteni, hogy a négy okon kívül nincs más ok, amikor e két további jelentkezőt sikeresen beillesztette a négy ok rendszerébe.

Úgy vélem, mindez nagyon hasonló ahhoz, amit Arisztotelész a *Metafizika* Alfa könyvének történeti áttekintésében (3–7. fejezet) általában, illetve specifikusan Anaxagorasz és Empedoklész princípiumainak esetében tesz. Miután a könyv első két fejezetében meghatározta, hogy a bölcsesség tárgyai a létezők elsődleges princípiumai és okai, rátér annak tárgyalására, hogy mik is ezek az elsődleges princípiumok és okok. Ehhez először is veszi a négy oktípus *Fizikában* kidolgozott rendszerét. Erre minden jogalapja megvan, hiszen ez az oksági elmélet már bizonyította erejét, mivel a természetfilozófián belül adekvát leírásokat szolgáltatott a jelenségek nagyon széles körére. Amikor most Arisztotelész a bölcsességet, a legmagasabb rendű tudást teszi meg tárgyául, és vizsgálni kezdi az annak sajátos tárgyait képező elsődleges princípiumokat és okokat, akkor nem kell mindent előlről kezdenie, hanem már támaszkodhat a *Fizika* alapvető eredményeire.<sup>16</sup> Teheti ezt annál is inkább, mert ezen a ponton a bölcsesség és a természetfilozófia viszonya még nem is került tisztázásra – akár még az is kiderülhet, hogy a természetfilozófia maga a bölcsesség. Ám miután kiindulópontként vette az okok *Fizikában* kidolgozott rendszerét, végigtekint azon elődei nézetein, akikről leginkább mondható, hogy a bölcsességgel foglalkoztak, és megvizsgálja, hogy az általuk említett princípiumok, illetve okok között van-e olyan, ami a négy ok *Fizikabeli* rendszerében további kiegészítést vagy módosítást tenne szükségessé. Azaz éppen úgy jár el, mint tette ezt a *Fizikában* a véletlen és a spontaneitás vizsgálatakor. A történeti áttekintés végeredménye is az lesz, mint a véletlen és a spontaneitás tárgyalásáé: nincs szükség további okfajták bevezetésére, mert a lehetséges további jelentkezők egytől egyig integrálhatók a rendszerbe.

Ehhez a konklúzióhoz természetesen szükség van egy jó adag értelmezői, elemzői munkára. Hiszen például ahhoz, hogy eldönthessük, vajon a véletlen és a spontaneitás a négytől alapvetően különböző okok-e, mindenképp meg kell értenünk, hogy mi is az a véletlen, és mi is az a spontaneitás. A már többször idézett ítélet, mely szerint „mindkettő a mozgás forrása oktípushoz tartozik”, közvetlenül függ a megelőző mondattól, amely szerint „[m]egmagyaráztuk te-

<sup>16</sup> A *Metafizika* Alfa könyve egyértelműen feltételezi a *Fizikát*, és több ponton explicit módon is utal rá. A harmadik fejezet elején éppen a négy ok bevezetését a következő megjegyzéssel zárja Arisztotelész: „Ezeket az okokat kielégítő módon tárgyaltuk a természetről szóló művünkben” (*en toisz peri phüszeszosz*, 984a32). Lásd még 985a12.

hát, hogy mi a véletlen, és mi a spontaneitás, és azt is, hogy miben különböznek egymástól”. Fontos látnunk viszont, hogy miközben Arisztotelész azon dolgozik, hogy e két jelenséget megértse és megkülönböztesse, már támaszkodhat a négy ok rendszerére a körkörösség legkisebb veszélye nélkül, hiszen a vizsgálat tárgya éppen az, hogy e kettő hogyan viszonyul ahhoz a négyhez.

A *Metafizikában*, az elődök vizsgálatánál egészen hasonló a helyzet. Ahhoz, hogy kiderüljön, vajon az általuk kidolgozott elméletek szükségessé teszik-e a négy ok elméletének módosítását, mindenekelőtt meg kell értenünk, hogy az elődök mit mondtak az okokról és a princípiumokról. Viszont amikor Arisztotelész azt próbálja megérteni, hogy az elődök mit is állítottak, illetve az általuk kidolgozott álláspontok miben és mennyiben különböznek egymástól, illetve hogy ezek az elméletek szükségessé teszik-e saját elmélete módosítását, akkor teljes joggal és a körkörösség legkisebb veszélye nélkül támaszkodhat a négy ok elméletére. Valójában nem is tehetne másként: amikor azt vizsgálja, hogy ezek az elméletek mit jelentenek a saját elméletére nézve, illetve ahhoz képest, akkor azokat nyilvánvaló módon a saját elméletének terminusaiban kell értelmeznie.

Talán nem is kell mondanom, hogy a történeti áttekintés ezen rekonstrukciójával egyszerűsre azzal a közkeletű nézettel is vitatkozom, amely szerint Arisztotelész illegitim módon, tendenciózusan és talán még rosszhiszeműen is félreértelmezi elődei nézeteit, rájuk erőltetve egy olyan fogalmi és értelmezési rendszert, amely e korábbi bölcselőktől teljességgel idegen volt.

A *Fizikában* a véletlen és a spontaneitás integrálása a négy ok elméletébe nem azt jelenti, hogy egyszerűen behelyezzük őket a négy doboz valamelyikébe. Azért nem tehetünk így, mert e kettő nem egyértelmű eset: néhány fontos megszorítást és pontosítást kell tennünk – például bevezetve az akcidentális okozás fogalmát –, és csak ezt követően sorolhatjuk őket a mozgató okhoz. Az elődök princípiumainak és okainak integrálása is csak hasonlóan fontos megszorítások után történhet: a korai bölcsék csak „az előzőekben mondtak szerint, és eddig a pontig jutottak” azon két ok vizsgálatában, amelyekről úgy tűnik, hogy valóban beszéltek. Ebben az esetben a megszorításokat és pontosításokat nem magának a dolognak a természete teszi szükségessé – nem az, ahogy a véletlen különbözik a mozgató okság standard eseteitől –, hanem az, *ahogyan* az elődök beszéltek ezekről az okokról és princípiumokról: pontatlanságaik, következetlenségeik, obskúrus, költőieskedő megfogalmazásaik. Az, hogy nem használták ki és nem tették explicitté saját felfedezéseiket, és tisztázatlanul hagyták fogalmaikat és magyarázataikat. Ha viszont annyi kiderül, hogy az általuk használt princípiumok és okok egyrészt nem mennek túl az általunk ajánlott négyen, pusztán zavarosabb megfogalmazásai e négynek, másrészt viszont valóban okok és princípiumok, akkor a legjobb, amit tehetünk, hogy a szükséges pontosításokkal és megszorításokkal hozzárendeljük őket az általunk ajánlott négyhez, hogy azután végül elmondhassuk:

Az erre a két okra vonatkozó vizsgálódások tehát az előzőekben, mint mondtuk, eddig jutottak (985b20–22).

## IRODALOM

- Bodnár István – Pierre Pellegrin 2006. Aristotle's Physics and Cosmology. In Mary Louise Gill – Pierre Pellegrin (szerk.) *A Companion to Ancient Philosophy* (Blackwell Companions to Philosophy). Oxford, Blackwell. 270–291.
- Broadie, Sarah 2009. The Ancient Greeks. In Helen Beebe – Christopher Hitchcock – Peter Menzies (szerk.) *The Oxford Handbook of Causation*. Oxford, Oxford University Press. 21–39.
- Burnyeat, Myles 2004. Introduction: Aristotle on the Foundations of Sublunary Physics. In Frans de Haas – Jaap Mansfeld (szerk.) *Aristotle's On Generation and Corruption I*. Oxford, Clarendon. 7–24.
- Dorion, Louis-André 2000. *Xénophon, Mémoires. Tome 1: Introduction générale*. Paris, Les Belles Lettres.
- Falcon, Andrea 2005. *Aristotle and the Science of Nature*. Cambridge, Cambridge University Press.
- Frede, Dorothea 1997. *Platon. Philebos*. Göttingen, Vandenhoeck & Ruprecht.
- Johnson, Monte Ransome 2005. *Aristotle on Teleology*. Oxford, Clarendon.
- Judson, Lindsay 1991. Chance and 'Always or for the Most Part'. In Lindsay Judson (szerk.) *Aristotle's Physics: A Collection of Essays*. Oxford, Clarendon Press. 73–100.
- Palmer, John 2009. *Parmenides and Presocratic Philosophy*. Oxford, Oxford University Press.
- Sedley, David 2007. *Creationism and Its Critics in Antiquity*. Berkeley – Los Angeles – London, University of California Press.

## Okság és morális felelősség Plótinosznál

### I. HELLENISZTIKUS ÉS CSÁSZÁRKORI ELŐZMÉNYEK

Az okság elve és a morális felelősség összeegyeztethetőségének kérdése – mely a mai filozófiában centrális probléma – az antikvitásban viszonylag későn, a hellénisztikus kor elején merült fel. A hellénisztikus és császárkori viták főszereplői a sztoikusok: provokatív tételük szerint „minden a fátum szerint történik”,<sup>1</sup> vagyis a világban minden, a legapróbb esemény is előre meghatározott, bekövetkezése elháríthatatlan. A sztoikus fátumelmélet kidolgozása Khrüszipposz (Kr. e. 280–206) nevéhez fűződik. Khrüszipposz a kétértékűség elvére támaszkodó logikai érvet is felhozott a determinizmus védelmében (Cicero: *A végzettről* 20–21), amelynek már Arisztotelésznél megtaláljuk előzményeit (*Herméneutika* 9). Fátumelméletének sajátága, hogy kauzális megfontolásokon nyugszik: „minden előzetes okok szerint történik” (Cicero: *A végzettről* 21). A sztoikus metafizika szerint a létezőket aktív és passzív princípium alkotja, ezért képesek hatást gyakorolni és hatást szenvedni. Oksági viszonyról akkor beszélhetünk, ha egy aktív ágensként ható test tevékenysége nyomán egy másik testről igazzá válik egy testetlen predikátum, melynek egy állapot vagy mozgás felel meg (például kés – hús – vágják). Khrüszipposz amellet érvelt, hogy mindennek van valami oka, és hogy különböző effektusokat csak különböző okok hozhatnak létre<sup>2</sup> – eszerint a világ egy adott állapota teljességgel meghatározza a következő állapotot. Már maga Khrüszipposz is szembenézett azzal a nehézséggel, hogy a determinizmus veszélyeztetni látszik a morális felelősséget,<sup>3</sup> ellenfelei pedig hevesen támadták ezen a ponton.<sup>4</sup>

A tanulmány az OTKA (K-75550) és a *Topoi Excellence Cluster* (Humboldt-Universität, Berlin) támogatásával készült.

<sup>1</sup> Diogenész Laertiosz 7.149; Diogenianosz, idézi Euszebiosz: *Praeparatio Evangelica* 6.8.1, 2, 6; Cicero: *A végzettről* 20. Az utóbbi mű fordítását tartalmazza Bugár–Lautner 2006.

<sup>2</sup> Plutarkhosz: *A sztoikusok ellentmondásai* 1045B–D. A szöveg fordítását tartalmazza Steiger 1983.

<sup>3</sup> Cicero: *A végzettről* 39–45; Gellius: *Attikai éjszakák* 7.2; Plutarkhosz: *A sztoikusok ellentmondásai* 1055F–1056D.

<sup>4</sup> A sztoikus determinizmus egyik korai kritikusa az akadémikus Karneadész, lásd Cicero: *A végzettről* 37.

A sztoikus fátumelmélet teleologikus színezetű: a világ eseményeinek lefolyása nem csupán megváltoztathatatlan, hanem egyben a legtökéletesebb tervet követi. A sztoikusok a fátumot értelemnek vagy oknak nevezik, valamint azonosítják a mindenség közös természetével, Zeusz akaratával, értelmével és a gondviseléssel. A rendező isteni princípium a lehető legjobb és legracionálisabb rendet hozza létre. A teleologikus kozmológia a sztoikusokat összeköti Platónnal és Arisztotelésszel, noha a platonikusok többsége és az arisztotelianusok vonakodnak az isteni gondviselést az egyedekre mint egyedekre kiterjeszteni. A világ isteni teremtését és rendjének célszerűségét az atomista hagyomány képviselői hevesen elutasítják.<sup>5</sup> E hagyományon belül is felmerül a determinizmus kérdése. Leukipposzt és Démokritoszt az utódok deterministának tartják (vö. Leukipposz fr. B2 DK). Epikurosz fellép „a természetfilozófusok szükség-szerűsége” ellen (*Levél Menoikeuszhoz* 133–134), jelentősen átalakítja az atomista fizikát, és érveket dolgoz ki a morális felelősség megalapozására (*A természetről* 34. könyv).

A császárkorban a peripatetikuskok a sztoikus fátumelmélettel szemben indeterminisztikus metafizika és kozmológia alapján érvelnek. Aphrodisziaszi Alexandrosz a 2–3. század fordulóján a fátumot az arisztotelészi természetfogalommal azonosítja. Mivel a természeti világban van kontingencia, a fátum nem áthághatatlan törvény, csupán tendencia. A dolgok „fátum szerinti”, vagyis a kérdéses szubsztanciák természetének megfelelő menete többnyire (*hósz epi to poliü*) érvényesül, ám megenged kivételeket. Alexandrosz a morális felelősség alapjává egy olyan képességet tesz, melynek révén nem csupán a külső körülményektől, hanem saját belső diszpozícióinktól is függetlenül dönthetünk. Meghatározása szerint rajtunk áll az, aminek az ellenkezőjére is képesek vagyunk.<sup>6</sup>

A szokásos rekonstrukció szerint a középső platonikusok hasonló módon kritizálják a sztoikus fátumelméletet.<sup>7</sup> A fátum nézetük szerint törvény, vagyis általános és hipotetikus jellegű: ha  $x$  esemény bekövetkezik, a következő események elkerülhetetlenek, azonban  $x$  maga nem a fátum rendelése, hanem a „rajtunk múló” dolgok körébe is tartozhat. Például ha Laiosz fiút nemz, akkor utóda kezétől kell pusztulnia, az előbbi azonban nem sorsszerű. A középső platonikus felfogásban a rajtunk múló cselekedetek a kontingens eseményeknek azon alosztályát alkotják, amelyek bekövetkezése egyenlő valószínűséggel bír ellentétük bekövetkezéssel, és valamilyen szenvedélyből vagy elhatározásból

<sup>5</sup> Aphrodisziaszi Alexandrosz: *A végeztéről* 12, 180, 4 sk. és 20 sk. A mű fordítását tartalmazza Bugár–Lautner 2006.

<sup>6</sup> Aphrodisziaszi Alexandrosz: *A végeztéről* 6 illetve 12.

<sup>7</sup> Az elmélet legjelentősebb kifejtése Pszeudo-Plutarkhosz: *A végeztéről*. További megfogalmazások: Nemesziosz: *Az ember természetéről* 34. 36–37 és 43; Alkinoosz: *Platón tanítása* 26; Calcidius: *Timaios-kommentár* 142–190. Pszeudo-Plutarkhosz művét magyar fordításban tartalmazza Steiger 1983, Nemesziosz munkáját Vidrányi 1994, Alkinoosz kompendiumát pedig Somos 2005.

fakadnak. A középső platonikus álláspontnak létezik egy újabb, alternatív értelmezése is.<sup>8</sup> Eszerint a középső platonikusok lényegében abban térnek el a sztoikusoktól, hogy míg a sztoikusok egyetlen kauzális forrást ismernek el a kozmoszban, és ezért nehezen tudják szétválasztani az emberi döntést és az isteni tervet, valamint mentesíteni az isteni gondviselést a rosszért való felelősség alól, ők e problémák megoldására két, egymástól független kauzális forrást tételeznek. A transzcendens, anyagtalan Isten és a kozmikus rendet megelőző anyag egyaránt „túllóg” a kozmoszon. Isten csupán a természeti törvényekért felelős, melyeknek az anyagra alkalmazásával a szublunáris világ mulékony egyedei létrejönnek, de nem felelős magukért az egyedekért, így az emberi jellemekért sem – ezek a véletlen művei. A cselekedetekért való felelősség ezért nem hárítható át a cselekvőről Istenre. A hipotetikus fátum tételében a feltételeket nem egyes események jelentik, hanem az események előzetes adottságai, beleértve a cselekvő egyének természetét. Ha ezek adottak, az eseménysor lefutása szükségszerű. A középső platonikus elmélet eszerint a kozmoszon belüli egyetemes kauzális meghatározottság kérdésében közelebb áll a sztoikus nézethez, mint az előbbi interpretáció szerint.

Plótinosz (Kr. u. 205–270) az oksági meghatározottság és a morális felelősség kérdéseiben elfoglalt álláspontját gyakran a peripatetikus és a standard módon értett középső platonikus tannal rokonítják.<sup>9</sup> A következőkben három

<sup>8</sup> Boys-Stones 2007; Sharples 2009.

<sup>9</sup> Plótinosz az okság és felelősség kérdésében elfoglalt álláspontjáról mindeddig nem bontakozott ki részletekbe menő, strukturált vita az irodalomban, főként kommentárokból vagy történeti áttekintésekből találunk idevágó megjegyzéseket. Plótinosz kozmológiájának determinista olvasata is felmerült. Amand (1945/1973. 155–163) szerint Plótinosz a sztoikus gondviseléstán nyomán elfogadja a fizikai világ egyetemes oksági meghatározottságát, ugyanakkor az emberi akarat szabadságát kétségbevonhatatlan ténynek tekinti; összegegyeztetetőségüket a gyakorlati cselekvés vonatkozásában nem képes megmutatni, a szabadságot a spiritualitás szférájára korlátozza. Graeser (1972. 112–125) elemzése hasonló nyomvonalon halad: az empirikus én alá van vetve az okságnak, ezen a szinten csak a cselekvés morális minősége függ tőlünk, metafizikai éntünk valódi szabadsága a fizikai meghatározottságoktól független *theóriában* van. Leroux (1996) hangsúlyozza, hogy Plótinosz számára az emberi szabadság nem mond ellent a szükségszerűségnek; tárgyalása szorosan követe Plótinosz megfogalmazásait, de nem tisztázza sem érveit, sem azok fogalmi előfeltételeit. Újabbban elterjedtebb az indeterminista értelmezés. White (1985. 215–245) amellett érvel, hogy Plótinosz végső soron indeterminista és inkompatibilista, és az emberi szabadságot, illetve felelősséget a racionális indokok (*reasons*) – a *theória* – birodalmába helyezi, melyet leválaszt a fizikai okok (*causes*) birodalmáról. White – aki Arisztotelész és a sztoikusok álláspontjáról kiváló, lényeglátó elemzést ad könyvében – kevésbé bocsátkozik bele Plótinosz tényleges érveinek elemzésébe. Brisson és Laurent (2002. 143–147) szerint Plótinosz *A végszetről* (III.1) szülő értekezésben középpont áll az epikureus indeterminizmus és a sztoikus determinizmus között, s a szublunáris szféra eseményeinek vonatkozásában megengedi a meghatározatlanságot – azonban *A gondviselésről* (III.2–3) írott munkájában a sztoikus determinizmust fogadja el. (Plótinosz műveinek rendjéről és idézésük szokásos módjáról lásd a következő jegyzetet.) A sztoikus determinizmus körüli hellénisztikus és császárkori viták fogalmi és argumentatív szerkezetének feltárásában mérföldkő Bobzien (1998) elemzése. Bobzien Plótinosz csupán futólag említi: egy lábjegyzetben további kommentár nélkül egytűvé sorolja a középső platonikusokkal és a keresztény

tézis mellett szeretnék érvelni. (1) Plótinosz felfogása a világ oksági folyamatairól közelebb áll a sztoikusokéhoz, mint feltételezni szokás: a kozmoszon belül az általános oksági determináció elvét vallja, és tagadja, hogy a jövő nyitott volna. (2) A morális felelősség kérdésében ennek megfelelően egyfajta kompatibilista álláspontot foglal el. (3) A morális felelősség kérdése és a szabadság, a normatív autonómia ideálja rendszerszerűen összekapcsolódik elgondolásában.

## II. A FIZIKAI VILÁG OKSÁGI RENDJE

Plótinosz egyik legkorábbi írása a Porphüriosz-féle kiadásban *A végzetről* címet kapta (III.1 [3]).<sup>10</sup> A kiinduló probléma az, hogy az oksági elv érvényes-e a valódi létezőkre, illetve a keletkező dolgokra, valamint hogy érvénye általános-e (III.1.1, 1–8). A fátum hagyományos problémája a kifejtés során kerül elő. A vizsgált kauzális elméletek értékelésének mindvégig eminens szempontja, hogy mennyiben hagynak helyet az emberi autonómiának. Ami a kauzális elvet illeti, Plótinosz álláspontja a következő. Az örök dolgok közül az „elsődlegeseket” (*próta*) nem lehet más okokra visszavezetni, míg mások czecktől függenek és bírják létüket (1, 8–11). Plótinosz nem mondja meg, hogy az örök, valódi létezők közül melyek minősülnek elsődlegesnek. A többes szám azt sejteti, hogy a szó tágan értelmezendő, az Egy/Jó – a modell egyetlen szigorú értelemben véve másra visszavezethetetlen entitása – mellett több intelligibilis létezőre alkalmazható. Az égitestek és a keletkező dolgok tekintetében viszont Plótinosz az oksági elv egyetemes érvénye mellett kötelezi el magát.<sup>11</sup> Ezt első körben azzal támasztja alá, hogy érveket hoz fel mind az atomok vagy egyéb testek ok nélküli mozgásával, mind pedig a lélekben szeszélyesen fellépő késztetéssel (*hormé empléktosz*) szemben (1, 13–24).<sup>12</sup>

A továbbiakban Plótinosz az általános kauzális elv értelmezését dolgozza ki. Vizsgálódása az érzékelhető dolgokra korlátozódik, pontosabban a szublunáris, keletkezésnek és pusztulásnak alávetett világ eseményeinek okaira (az égi szféra csak mint lehetséges kauzális tényező szerepel). Kezdetben az oksági relációknak csak egy bizonyos típusát veszi figyelembe. Az értelemmel felfogható világ princípiumai közül monista rendszerében csak akkor nevezhet többet is

szervezőkkel, akik előkészítették a talajt Alexandrosz indeterminista felelősségfelfogása számára (345. 43. jegyzet).

<sup>10</sup> Plótinosz írásait tanítványa kilenc-kilenc értekezést tartalmazó csoportokba, *enneaszokba* rendezte. E megjelölést címként is használják („Plótinosz *Enneaszai*”). A szokásos idézési mód szerint római szám jelöli az *enneasz*, arab szám az értekezés számát, []-ben megadható az adott mű kronológiai rend szerinti sorszám, végül ehhez csatlakozhat a fejezet- és a sorszám.

<sup>11</sup> Vö. Platón: *Timaios* 28a.

<sup>12</sup> Khrüszipposz hasonlóképpen kritizálja a lélek váratlan mozgásainak tézisért, lásd Plutarkhosz: *A sztoikusok ellentmondásai* 23–24.



„elsőnek”, ha a vertikális metafizikai függéstől eltekint. A vertikális oksági relációk alapvető fontosságúak a három hiposztázis közötti viszonyoknak, valamint az értelemmel felfogható valóság és az érzékelhető világ relációjának a leírásában.<sup>13</sup> Jelen fejtegetésben azonban Plótinosz az oksági viszonyok szokványosabb értelmezéséből indul ki. Eszerint az okok és az okozatok ugyanazon a síkon helyezkednek el: a fizikai világban, melybe bele kell értenünk a lélekkel bíró, sőt a racionális élőlényeket is.

Az egyetemes oksági elv védelmét követően Plótinosz először a „közvetlen okokat” (*aitiai proszekheisz*) megjelölő elemzésre mutat be példákat (1, 24–36).<sup>14</sup> A magyarázandó tények példajaként cselekvések, szubsztanciák keletkezése és egyéb események szerepelnek. A példák elrendezése hagyományos platóni-arisztotelészi sémát követ,<sup>15</sup> az elemzésbe azonban a sztoikus cselekvépszichológia és okságelmélet elemei is beszüremkednek. (1) *Tudatos emberi ágencia*. A piacra menetel oka az a vélekedés, hogy szükséges valakivel találkozunk vagy a kölesönt megkérünk. A példa általánosításában Plótinosz nem magát a cselekvést, hanem a rá irányuló választást és késztetést (*heleszthai, buleszthai*) okolja meg azzal, hogy az illetőnek az adott cselekvés tűnt jónak. A cselekvés szükségességét magában foglaló vélekedés vagy képzet (*oiéthénai, phanénai*) mint a cselekvés kiindulópontja, valamint az okozati oldalon a cselekvés pszichikus oldalának hangsúlyozása kölesönzés a sztoikusoktól.<sup>16</sup> (2) *Mesterség*. A gyógyulás oka – kifogástalan arisztotelészi módon – az orvosi mesterség vagy az orvos. (3) *Véletlen egybeesés*. A meggazdagodás oka a kincstalálás, valakinek az adománya, vagy a pénzkeresés valamilyen mesterségből. A kincstalálás standard példa a véletlen eseményre,<sup>17</sup> a pénzkereset pedig csupán járulékosan tartozik hozzá például az orvosi mesterséghez. Plótinosz elkerüli a véletlen (*tükhé*) említését, és egyáltalán nem állítja, hogy a véletlen a természeti világban ható okok egyike. Ehelyett a véletlen standard példáját megpróbálja hatástalanítani azzal, hogy rámutat: ennek is meg tudjuk jelölni valamilyen okát, ezért nem veszélyezteti az okság egyetemes elvét. (4) *Természet*. A gyermek születésének oka az apa, mellette a táplálék, a bőséges sperma és a szülésre alkalmas nő csupán közreha-

<sup>13</sup> A vertikális kauzális viszonyokat Plótinosz a kettős *energeia* modelljével értelmezi, ehhez lásd Emilsson 2007. 22–68.

<sup>14</sup> A kifejezés előfordul Alexandrosz Arisztotelész-kommentárjaiban, lásd pl. *Metafizika-kommentár* 19. 29.

<sup>15</sup> Plátón: *Törvények* 888–889e: *phüszisz, tükhé, ananké* vs. *nusz, theosz, tekhné*; *A szofista* 265c; vö. még *Philebosz* 28d, 30c. Arisztotelész: *Nikomakhoszi etika* 1112a33: *phüszisz, tükhé, ananké, nusz*; Theophrasztosz, *Sztobaiosz*: *F.c.* 1.6.17c: *phüszisz, tükhé, ananké, proaireszisz*.

<sup>16</sup> Az előbbi a sztoikus *phantasztia hormétiké* körülírásának tűnik (vö. *Sztobaiosz*: *F.c.* 2.7.9), esetleg beleértve a sztoikus elméletben kulcsfontosságú jóváhagyás (*szünkatatheszisz*) mozzanatát is; az utóbbi a cselekvésnek a *hégemonikon* mozgásával való khrüszippuszi azonosítására emlékeztet (Seneca: *Erkölési levelek* 113.23). Az imént Plótinosz még arisztoteléianus módon a törekvés (racionális vagy irracionális vágy) tárgyát jelölte meg a cselekvés mozgatójaként (1, 22–24).

<sup>17</sup> Arisztotelész: *Nikomakhoszi etika* 1112a27, vö. *Metafizika* 1025a16 sk.

tó oknak (*szünergon*) tekinthető.<sup>18</sup> Az ilyen folyamatok általánosságban az illető dolog természetére vezethetők vissza.

Plótinosz nem elégszik meg a közvetlen okok feltárásával, hanem „az első, ezeken túli okok” (*ta próta kai... epekeina aitia*, 2, 3 sk.) felderítését szorgalmazza. Mi okozza azt, hogy ugyanazon körülmények között más és más cselekvők másként cselekszenek? Mi az oka a szokások, jellemek és sorsok (*tropoi, éthé, tükhai*, 2, 8) különbségének? A cselekvő adottságaira végső magyarázatot javasló elméletek közül Plótinosz négyet vizsgál meg és vet el (3–7. fejezet): (1) minden dolog, köztük a cselekvők tulajdonságai is a testi konstituensek (atomok vagy elemek) tulajdonságaira vezethetők vissza; (2) a világ minden eseménye egyetlen ágensre, a világlélekre vezethető vissza, az egyes élőlények – köztük az emberek – pusztán organikus részei a kozmikus élőlénynek; (3) a szublnáris világ eseményeit az égitestek határozzák meg; (4) a világ eseményei kauzális láncolatot alkotnak, az események elkerülhetetlenül bekövetkeznek, ez azonban nem veszélyezteti az emberi autonómiát.<sup>19</sup>

Plótinosz cáfolatait itt nincs lehetőségünk elemezni. Vegyük viszont szemügyre azokat a követelményeket, melyeket a korábbi elméletek aporetikus vizsgálata után egy adekvát okságmélettel szemben állít!

Milyen más ok van tehát még, mely hozzájárul ezekhez [ti. a sztoikusok által elismert testi okokhoz], és [a] semmit nem hagy ok nélkül, [b] megőrzi az események egymásra következését és rendjét, [c] megengedi, hogy mi is legyünk valamik, [d] és nem szünteti meg az előrejelzést és a jóslást? Nos, a lelket kell bevezetnünk a létezők közé másik princípiumként, és így kell egybeszólnünk minden dologot – nem csupán a mindenség lelkét, hanem vele együtt az egyes élőlényekét is, hiszen ez nem jelentéktelen princípium, s nem magvakból jön létre ő is, mint a többi dolog, hanem elsődlegesen ható ok (III.1.8, 1–8).

<sup>18</sup> A *szünergon* sztoikus eredetű terminus (Sextus Empiricus: *A pürrhónizmus alapvonalai* 3.15), jóllehet platóni előzményei is vannak. A szöveg fordítását tartalmazza Kendeffy 1998.

<sup>19</sup> Az (1) az epikureus atomizmus és közelebbről meg nem határozott egyéb korporealista elméletek; a (2) nem azonosítható pontosan, talán sztoicizáló platonikus nézet; a (3) a császárkorban népszerű asztrális determinizmus, melyet Plótinosz a II.3 [51]-ben kritizál részletesen; a (4) a sztoikus kompatibilizmus. A kritizált nézetek közül hiányzik a peripatetikus Alexandrosz nézete, akinek *A végszetről* írott művét Plótinosz nagy valószínűséggel ismerte. A kritika két okból maradhat el: (a) Plótinosz nagyjában-egészében elfogadja a természeti folyamatoknak a „közvetlen okokat” megadó elméletét, de felületesnek tartja (vö. III.1.2, 1–4), ezért nem ennek bírálatából bontja ki nézetét; (b) Alexandrosz felfogása a természeti okságról nem jelent fenyegetést az emberi autonómiára nézve, melynek megalapozására Plótinosz vállalkozik.

Plótinosz fenn kívánja tartani az egyetemes oksági elvet, mely mellett az 1. fejezetben is érvel. A okok szövedéke, mely egybetartja a mindenséget, egyszersmind értelmes, célszerű rend.<sup>20</sup> Az emberi autonómia mind az eddigi aporetikus vizsgálódásban, mind az innen kezdődő konstruktív fejtegetésben elsőrendű szempont. A jóslás az antikvitásban széles körben elfogadott, az állami életben is szerepet játszó tudomány volt, melynek olykor még a sztoikusok kritikusan is megpróbáltak helyet találni. Plótinosz e követelményeknek két stratégiai húzással kíván eleget tenni. (A) A testi, fizikai okok mellett az okoknak egy másik típusát is bevezeti: a lelkeket, melyek az okság elsődleges forrásainak minősülnek. (B) Egymástól független kauzális princípiumok sokaságát ismeri el. A világ eseményei nem vezethetők vissza egyetlen pszichikus princípiumra, a világlélekre, az egyéni lelkek nem csupán a testek vonatkozásában autonóm ágensek, hanem a világlélek vonatkozásában is.

Érdemes megvizsgálunk az egyetemes kauzális elv érvényesülésének további fejtegetését.

Mindaz tehát, ami elhatározás (*proaireszisz*) és véletlenek (*tükhai*) keveredésével jön létre, szükségszerű – hiszen mi másra gondolhatnánk még ezeiken túl? Ha minden okot számításba veszünk, minden esemény feltétlenül bekövetkezik (*panta pantósz ginetai*); a külső tényezők között szerepel az is, ha esetleg valami az égi körforgásból is hozzájuk járul (III.1.9, 1–4).

Észerint *a kozmoszon belül* a kiinduló feltételek okságilag teljesen meghatározzák az eseményeket. A kiinduló feltételek között szerepel az emberi ágencia is (erre utal az „elhatározás” terminus). A „véletlenek” (*tükhai*) kifejezés jelen érvben, mely az egyetemes kauzális meghatározottság mellett szól, nem utalhat meghatározatlan, bizonytalan tényezőkre, amelyeknek köszönhetően a kérdéses események többféleképpen is lejátszódhatnak, hanem a kérdéses cselekvés ágensének természetéhez képest *külsődleges* tényezőket takar (amilyen például a saját test, az élőlények, vagy akár az égitestek). Jelen megfogalmazásnak (9, 2–3) pontos megfelelőjét találjuk a sztoikusok elméletének kritikai tárgyalásában:<sup>21</sup>

[E nézethez] tartozik, hogy minden feltétlen szükségszerűséggel bekövetkezik (*tén pantósz pantón anankén*), és ha minden okot számításba veszünk, nem lehetséges, hogy ne következzenek be minden egyes esemény (III.1.7, 8–10).

<sup>20</sup> Az elmélet e teleologikus oldalához vö. IV.4.33 és 35: a véletlennek nincs helye a mindenség életében, mely egyetlen harmóniát és rendet alkot.

<sup>21</sup> Vö. III.4.5, 13 sk.: *hai... exóthen tükhai tén proaireszin uk ekkibadzusin*. Plótinosz másutt hasonló értelemben használja a *szüntükhiai* (véletlen körülmények) szót, és hangsúlyozza, hogy ezek a lélek számára külsődleges tényezők szigorú oksági rendbe illeszkednek (III.3.2, 1–3).

Ez első olvasásra szemrehányásnak hangzik, a 9. fejezet elejének párhuzama azonban azt sejteti, hogy Plótinosz valamilyen formában saját elméletébe is beépíti a kozmoszon belüli determinizmust.<sup>22</sup> Vizsgáljuk meg, hogy vajon kiállja-e ez az értelmezés a szövegekkel való szembesítés próbáját.

Kulcskérdés, hogy miként kívánja Plótinosz használni az emberi ágenciát a kozmosz kauzális rendjének magyarázatában. Elvileg elképzelhető, hogy épp ezen a ponton akarja fellazítani a világfolyamat determinisztikus rendjét azon az alapon, hogy az elhatározás okként szolgál a további események láncolatában, ám nincs olyan ok, mely őt magát előzetesen meghatározná. Ennek azonnal el-lene szól, hogy az emberi ágenciát („ami tőlünk függ”) teljesen másként közelíti meg, mint akár Alexandrosz, akár a középső platonikusok: míg az utóbbiaknál a tőlünk függő dolgok köre a lehetséges (*dünaton*), közelebbről a kontingens (*endekhomenon*) egy alosztálya, nála e terminusok szakkifejezésként elő sem fordulnak.<sup>23</sup> További ellenérv, hogy Plótinosz nem tételez diszkontinuitást sem a jellem és a cselekvés, sem a lélek előtörténete és a jellem között.<sup>24</sup> Még mindig nyitva áll azonban a lehetőség, hogy ezek a kauzális kapcsolatok nem jelentenek teljes meghatározottságot.<sup>25</sup> Plótinosz elfoglalhatná azt az álláspontot, hogy a cselekvés oka a cselekvő döntése, mely nem független jellemétől, előlétele pedig vitathatatlanul rajta hagyta a nyomát jellemén, ez azonban nem zárja ki, hogy az említett kauzális tényezők nem korlátozzák lehetőségeit egyetlenegyre, és a döntés pillanatában valódi alternatívák között választhat. Ha ez így volna, a lelkek kauzális szerepének köszönhetően a világ eseményei nem alkothatnának egyetlen, elágazások nélküli, lineáris oksági sort, hanem többféleképpen is alakulhatnának. De vajon nyitott-e a jövő Plótinosz felfogásában?

Több jel utal arra a szövegekben, hogy e kérdést nemlegesen kell megválaszolniunk. A III.1 első fejezete kizárja az ok nélküli mozgásokat mind a fizikai testek, mind a lélek esetében. Persze ez önmagában nem sokat mondana, hiszen az ok nélküli mozgás hírhedt epikureus tanát az indeterminista Alexandrosz is kínosan kerüli. Érdekesebb, hogy az atomizmus kritikája szerint jóslás – mind tudományos, mind isteni inspirációnak köszönhető válfaját tekintve – csak akkor lehetséges, ha a jövő meghatározott (*dei... hórízsmenon to mellon einai*, 3, 16 sk.). Plótinosz érvelése szerint a jóslás gyakorlatának megvannak a maga ontológiai-kauzális előfeltételei, melyeket az atomista nézetrendszer nem biztosít: a rend (*taxisz*) és meghatározottság. Kérdés persze, hogy mi mindenre terjed ki a rend, vajon univerzális érvényűnek kell-e gondolnunk. Szövegrészünkből szi-

<sup>22</sup> Az okok – az emberi ágencia és a külső körülmények – konjunkciója *A gondviselésről* írott értekezésben is feltétlenül meghatározza a kimenetelt (III.2.10, 11 sk.: *hé de ananké uk exóthen, all' hoti pantósz*).

<sup>23</sup> Lásd Sleeman–Pollet 1980, *dünaton*, illetve *endekheszthai* címszó.

<sup>24</sup> Erre a következő szakaszban térek ki bővebben.

<sup>25</sup> Cicero: *A végzet* 7–10 hasonló nézet alapján kritizálja a khrüszipposzi determinizmust; vö. uo. 23–25.

gorúan véve csak annyi következik, hogy sikeres jóslás esetén a bekövetkezendő események előre meghatározottak a jóslás időpontjában.

A kérdés szempontjából fontos A *gondviselésről* (III.2–3 [47–48]) írott értekezés. Plótinosz fő célja itt annak igazolása, hogy a világ providenciális rendjét nem kérdőjelezi meg a benne tapasztalható rossz, hogy azért nem az isteni princípiumokat terheli a felelősség. E kései mű okságfelfogása nézetem szerint nem különbözik lényegesen a III.1-ben kifejtett nézettől. Plótinosz ebben az értekezésben visszatér a jóslás kérdésére:

Mínthogy a mindenség élőlény, az, aki a benne lejátszódó történéseket szemléli, egyszersmind forrásaikat is szemléli, és a gondviselést, amely megjelenik rajta; ez utóbbi mindenre kiterjed, azokra a dolgokra is, amelyek keletkeznek: ezek az élőlények, cselekedeteik, valamint kevert diszpozícióik, amelyek az ész és a szükségyszerűség vegyülei [...] és ő [tudniillik a jós] nem tudja szétválasztani egyfelől a gondviselést és ami a gondviselés szerint van, másfelől pedig azt, ami alapul szolgál számára, ti. hogy mit ad hozzá ahhoz, ami belőle ered. [...] Mert hiszen nem a jós dolga megmondani a „miért”-et, hanem csak a „hogyan”-ot, és művészete a természet betűinek olvasása, melyek megmutatják a rendet, és sehol nem hanyatlanak a rendezetlenségbe – ehelyett az égi körforgás tanúskodik arról és napvilágra hozza, hogy milyen minden egyes ember, és milyen vonásokkal bír (*hoiosz hekasztosz kai hosza*), még mielőtt mindez saját tetteikből megmutatkozna (III.3.6, 8–22).

Ebben a szövegrészben megvan, amit A *végzetről* szóló értekezésből hiányoltunk. A jóslásról világosan kiderül, hogy univerzális, bármely emberi cselekedetre vonatkozhat, nem csupán a gondviselés racionális tervét előmozdító tettek, hanem a rossz cselekedetekre is (III.3.6, 1), amelyek összeszövődnek az előbbiekkel. Ennek az az alapja, hogy a gondviselés rendje, amely összhangba hozza a világban ható autonóm ágensek működéseit, szintén minden részletre kiterjed.<sup>26</sup> A nyitott jövő koncepciója tehát határozottan kizárható. A kozmikus eseménysor nem játszódhat le másként, mint ahogyan ténylegesen lejátszódik, vagyis a lehető legracionálisabb módon. E rend belátható az isteni rendező, a gondviselés szempontjából, sőt a jós is képes lehet a jövőbeli történések előrejelzésére.

<sup>26</sup> Alexandrosz szerint a gondviselés a szublanáris szféra egyedeire nem terjed ki (A *gondviselésről* 13, 14–17). A középső platonikusok standard tana hasonló lehetett: az érzékelhető világra vonatkozó másodlagos gondviselés, mely a fátummal azonos, *általános* törvény; a személyes *örzőszellemek* gyakorolta „harmadlagos gondviselés” (Pszeudo-Plutarkhosz: A *végzetről*<sup>9</sup>) vélhetőleg engedmény a standard elmélethez képest. Plótinosz – a sztoikusok álláspontját követve – ezzel szemben ismételten hangsúlyozza, hogy a gondviselés minden részletet magában foglal (III.2.6, 18–22; 7, 29–40; 13, 18–27).

Hasonló implikációkkal bír az értekezés két hasonlata. Plótinosz az emberi életet színjátékhoz hasonlítja, melyben a szerzőnek a gondviselés, a színésznek a lélek, a szerepnek pedig a konkrét személy felel meg. A szerző/rendező a színészre már meglévő adottságai szerint osztja rá a megfelelő szerepet és állítja a megfelelő helyre, de az már a színészen múlik, hogy jól vagy rosszul játszsza-e szerepét; ettől függ viszont, hogy legközelebb – a következő fellépésen, illetve a következő megtestesülés alkalmával – milyen szerepben találja magát (III.2.17–18). Szempontunkból fontos, hogy a kozmikus drámában a színész a szerzőtől kapja szövegét,<sup>27</sup> és hogy *nem tud változtatni a cselekményen*,<sup>28</sup> akár hangjának és gesztusainak szépségével, akár hangjának hitványságával járul hozzá a darabhoz. Az ember cselekedetei meg vannak szabva, lényegében nem változtathatók meg („nem maga írja a szövegét”), s a szerző, a gondviselés perspektívájából belátható összes jövőbeli cselekedete.<sup>29</sup> A másik hasonlatban a gondviselés megfelelője a hadvezér, aki előre látja a cselekvéseket és szenvedéseket, hogy milyen kellékekre – élelemre, itatra, fegyverre és hadigépekre – van szükség a hadviseléshez, s mindazt, ami ezek összeshívódésából adódik, hogy az eredményt is jól rendezhesse el; ráadásul a gondviselés olyan hadvezérhez hasonlít, aki nem csupán saját seregét, hanem az ellenfélét is irányítja (III.3.2, 3–15). E hasonlat szerint a katonák autonóm cselekvőkként működnek együtt a hadvezérrel, ő azonban végtelen bölcsességében minden autonóm tettetét át-látja és a legszebb rendbe foglalja:

Mert még ha azt mondanád is: „én képes vagyok ezt vagy amazt választani (*heleszthai*)”, amit választani fogsz, benne foglaltatik a rendben, hiszen a te cselekvésed nem epizódszerű a mindenség szempontjából, hanem bele vagy számítva abba mint ilyen és ilyen ember (*ho toioszde*) (III.3.3, 1–3).

<sup>27</sup> III.2.17, 20 sk. és 29: *tusz logosz*.

<sup>28</sup> III.2.17, 44 sk.: *uk epoiészen... to drama heteron hoion én*.

<sup>29</sup> A hasonlat további fejtegetésében Plótinosz még messzebb megy a determinizmus irányában: „Semmi szükség arra, hogy olyan színészeket léptessünk fel, akik valami egyebet mondanak, nem a költő szavait, mintha az utóbbi által írt színdarab tökéletlen volna, és a költő a darab közepén üres helyeket hagyott volna, ezért maguk pótolnák ki, ami hiányzik – hiszen így a színészek nem színészek, hanem a szerző részei lesznek, és pedig egy olyan szerzőé, aki előre tudja, mit fognak mondani, hogy ily módon képes legyen egybeszőni a darab hátralévő részét is [...] mi akadály, hogy a színészek tetteit – ahogyan ott a színdarabnak – a mindenség formáló elvének (*logosz*) részeivé tegyük, és azt is, hogy jól vagy rosszul játszanak, oly módon, hogy ezt minden egyes színész magától e formáló elvtől kapja” (III.2.18, 7–13 és 21–25). Úgy tűnik, Plótinosz itt némiképp túllő a célon, hiszen az értekezés koncepciója szerint emberi autonómia nélkül nem létezhetne gondviselés sem (III.2.9, 1–4). A fejtegetés közvetlen folytatásában pontosítja álláspontját: az egyetemes *logosz* a hitvány és derék tetteket részekként tartalmazza, de nem ez hozza létre őket, csupán velük együtt áll fenn (III.3.1, 1–4).

A cselekvő egyéni karaktere és egyes választásai eleve bele vannak kalkulálva az isteni tervbe, ő nem képes változtatni az események lefutásán.<sup>30</sup>

Összegezzük eredményeinket. Plótinosz, bár nem tüntet determinizmusával, mint a sztoikusok, a kozmoszon belüli események vonatkozásában az övékhez igen hasonló nézetet fogad el. A számításba veendő okokból az események szükségszerűen következnek, a világfolyamat nem megváltoztatható, a jövő előre meghatározott, ráadásul a jóslás révén az emberi megismerés számára is hozzáférhető. A fizikai okságnak („természet”, „fátum”) azonban – mely a sztoikusoknál az okság egyetlen fajtája, és a fátum/providencia rendjének hordozója – Plótinosz igen korlátozott szerepet tulajdonít, és mellette testetlen okokat is bevezet. További fontos különbség, hogy míg Plótinosz érvelésében kulcsszerepe van az egymástól független kauzális források tételezésének, a sztoikusok elméletében ez korántsem ilyen világos.<sup>31</sup> Plótinosz okságelméletét a nem fizikai okok tételezése miatt nem lehet közvetlenül összehasonlítani a modern determinista teóriákkal, hiszen az utóbbiak abból indulnak ki, hogy a világ kauzálisan zárt, nincsen benne hely nem fizikai oksági tényezőknek. A „determinizmus” címkével azért is óvatosan kell bánnunk, mert Plótinosz elméletében a lelkek, melyek a világ eseményeinek kauzális forrásai – jóllehet nem lehetnek másmilyenek, mint amilyenek – nem, vagy legalábbis nem csupán kívülről meghatározottak okságilag.<sup>32</sup> Mindamellent a *kozmoszon belüli események* vonatkozásában és az immateriális okoknak is helyt adva, kiterjesztett értelemben tulajdoníthatjuk Plótinosznak a determinizmus tételét. Ezzel az okságelmélettel Plótinosz a morális felelősség kérdésében csakis a kompatibilizmus valamilyen változatát képviselheti.

### III. MORÁLIS FELELŐSSÉG: AUTONÓMIA ÉS ÖNKONSTITÚCIÓ

A morális felelősség alapjának kézenfekvő azt tekinteni, hogy az adott cselekvő az adott külső körülmények és belső diszpozíciók mellett cselekedhet-e másképp („a másként cselekvés szabadsága”). Ez a kritérium az antikvitásban meg lehetőségen későn, de még Plótinosz előtt – a peripatetikus Alexandrossznál – merült fel. A hagyományos felfogás szerint a felelősség a cselekvő autonómiájában gyökerezik, vagyis abban, hogy egy adott cselekedetért a cselekvő felelős kauzálisan, és nem valamilyen más, külső tényező.<sup>33</sup>

<sup>30</sup> Sharples (1994) rámutat, hogy Plótinosz – Alexandrosszal szemben – kiterjeszti a gondviselést az egyedi dolgokra, és utal az ebből adódó episztemológiai nehézségekre.

<sup>31</sup> Plótinosz azt veti a szemükre, hogy mivel mindent egyetlen princípiumból származtatnak, az emberi cselekvőknek csupán instrumentális szerepet hagynak (III.1.7, 12 sk.).

<sup>32</sup> Ez utóbbi kérdéssel a következő szakasz végén foglalkozom.

<sup>33</sup> A morális felelősség e két koncepciójához lásd Bobzien 1998. 276–279.

A választást (*haireszisz*, *proaireszisz*) Plótinosz több helyen említi, azonban világosan kiderül, hogy felfogásában nem választásunk meghatározatlan, nyitott volta biztosítja felelősségünket. Az emberi cselekvés Plótinosz felfogásában mindig valamilyen célra – a vágy vagy a racionális akarat tárgyára – irányul, a készítés nem léphet fel spontán módon, ok nélkül a lélekben, mert az ok nélküli és akaratlan mozgások nagyobb kényszert jelentenének, mint az efféle meghatározottság (III.1, 18–22) – nyilván a választás sem működhet ezen a módon. Plótinosz abban az értelemben sem tekinti szabadnak a választást, mint Alexandrosz, akinek felfogásában ellentétes lehetőségek között dönthetünk, belső diszpozícióinktól is függetlenül.<sup>34</sup> Az egyes emberek Plótinosz szerint (egyedi) természetüknek megfelelően végzik cselekvéseiket, mást és mást tesznek egyéni különbségeik szerint, és tetteik, életük morális minőségét is természetük határozza meg (III.3.1, 24–27). Legutóbbi idézetünk implikálja, hogy választásaink egyedi jellegzetességeinket követik (III.3.3, 1–3).

A „választás” Plótinosznál nem csupán egyes cselekvésekkel kapcsolatban kerül elő. Platón *Államának* zárómítosza (617d skk.) szerint a lélek a születés előtt maga választja életformáját, ezért ő viseli a felelősséget, és nem az istenség felelős további sorsáért. Plótinosz ennek nyomán gyakran beszél „az életformák választásáról” (*haireszisz bión*).<sup>35</sup> Íme egy példa:

De ha [a lélek] ott [ti. a túlvilágon] választja meg daimónját és ha ott választja meg életformáját, hogyan vagyunk bárminek is urai? Talán az a helyzet, hogy az ottani választás, amiről [Platón] beszél, rejtvény formájában a lélek általános, mindent átfogó irányultságára és beállítottságára (*proaireszisz kai diatheszisz katholu kai pantakhu*) utal. De ha a lélek beállítottsága a döntő, és az a rész érvényesül benne, amely az előző életekben átélt tapasztalatok nyomán leginkább kéznél van, akkor nem a test az oka semmi rossznak számára; mert ha a lélek jellege megelőzi a testet, és az a jellemünk, amit választottunk [...], akkor derék, illetve hitvány sem itt [az érzékelhető világban] lesz az ember (III.4.5, 1–11).

<sup>34</sup> Az autonóm mozgással bíró dolgok olykor a jó, olykor viszont a rossz felé hajlanak (III.2.4, 36–44). Ezt hajlamosak lehetünk az indeterminisztikus választási képességre való utalásként olvasni. Plótinosz valójában nem erről beszél: a rossz felé fordulás vagy az értelem újbóli megtisztulása a karaktert alakító cselekedeteken keresztül történik az egymást követő megtestesülések során (III.3.4, 34–37). Más összefüggésben Plótinosz megengedni látszik az elmentétekre való képességet, amellyel Alexandrosz meghatározza az autonómiát, de ő itt sem ezt tekinti az autonómia meghatározásának (VI.8.21, 3–7).

<sup>35</sup> A fontosabb helyek: II.3.9 és 14–15; III.2.7, III.4.3 és 5–6, IV.3.12–14. Phillips (1995) helyesen hívja fel a figyelmet arra, hogy Plótinosz *proaireszisz*-fogalma különbözik Arisztotelész értelmezésétől.



A „választás” nem egyedi cselekvési szituációkban hozott döntés, hanem a lélek mitikus-metafizikai előtörténetének mozzanata, fogalmi nyelven szólva pedig a lélek diszpozíciója, irányultsága, melynek egyes elemei aktualizálódnak korábbi tettei nyomán, s ezzel nyeri el egyedi profilját. Egyéni karakterünket, melyből életünk menete és sorsunk következik, eszerint nem kívülről kényszerítik ránk, hanem saját magunk is részt veszünk meghatározásában. Tetteinkért azért vagyunk felelősek, mert jellemünkért felelősek vagyunk. Plótinosz a vizsgált kontextusban összeköti a választást és a felelősséget, de egészen másként, mint Alexandrosz.

Mindebből azt a következtetést vonhatjuk le, hogy Plótinosznak esze ágában sincs elvágni vagy lazítani a köteléket jellem és választás között, hanem a morális felelősség másik, hagyományosabb megközelítését képviseli: tetteinkért és választásainkért azért vagyunk felelősek, mert nem valamilyen *más, külső* tényezőkön múlnak, hanem a *mi* belső diszpozícióinkat, karakterünket tükrözik. Erről árulkodik, hogy Plótinosz elégségesnek tartja a felelősséghez, ha ugyanazon helyzetben *egy másik* cselekvő másképp cselekedne: Helené szépsége másképp hatott Pariszra, és másképp Idomeneuszra, aki szintén gyakran megfordult Menelaosz házánál, mégsem jutott eszébe megszőkötetni annak feleségét (III.3.5, 40–43).

A morális felelősség megalapozásának kulcsa Plótinosz okságelméletében az, hogy több, egymástól független kauzális forrás működését tételezi a kozmoszban. Ebben az összefüggésben az egyéni lelkek „elsődleges létezőnek” (*próton*) minősülnek, vagyis olyan tényezőknek, amelyeket okságilag nem határoz meg más, ők viszont „elsődlegesen ható okok” (*próturgoi aitiai*, III.1.8, 8), oksági folyamatok kiindulópontjai,<sup>36</sup> ezért lehetnek a morális felelősség alanyai.

Az egyéni lelkek kauzális függetlenségének tétele korántsem magától értendő, hiszen a Lélek a *harmadik* a hiposztázisok sorában, és az egyéni lelkek e hiposztázison belül sem a legmagasabb rendűek. Nem *ad hoc* hipotézisről van szó: Plótinosz standard tanításában két megfontolás is alátámasztja e tételt. Ezek egyike a lélek testetlensége.<sup>37</sup> Szigorú értelemben a lélek mint olyan nem szenvedhet hatást a testtől, ezért kauzálisan nem határozhatják meg fizikai tényezők. A másik a Lélek hiposztázisának belső szerkezetére vonatkozó elmélet:

Amennyiben elfogadják, hogy a mindenség lelke és az egyes lelkek ugyanazokkal a valóságokkal érintkeznek, ezzel elfogadják, hogy a kétfajta lélek hasonló természetű (*homoeidész*), vagyis egyazon nembe sorolják őket, ami kizárja, hogy egyik a másik része lehetne. Helyesebb volna azt mondaniuk,

<sup>36</sup> Az „elsődlegesen ható” kifejezést Platón is használja a lélek mozgásaira, amikor a léleknek a testtel szembeni oksági elsődlegességét bizonyítja (*Törvények* 897a4).

<sup>37</sup> Ezt a tételt *A lélek halhatatlanságáról* írott korai munkában részletes érveléssel támasztja alá (IV.7 [2] 2–8<sup>b</sup>).

hogy egyugyanazon lélekről van szó, és mindegyik lélek egész. Ha viszont egy lélekről beszélnek, ezáltal a lelkeket valami másra vezetik vissza, mely már nem tartozik ehhez vagy ahhoz, nem lelke semminek, sem a világnak, sem bármi másnak, hanem a világ és minden más eleven lény lelkének alkotója (IV.3.2, 1–8. ford. Horváth Judit – Perczel István).

Az emberi lélek nem része a világléleknek, nem az „hozza létre”, hanem mindketten egy további, közös forrástól, az egyetemes Lélektől származnak, mely már nem tartozik semmilyen testhez (IV.3.1–8; IV.9.4–5). Ezen elmélet szerint a világlélek és az egyéni lélek között nincsen hierarchia, lényegében egyenrangúak. Az emberi léleknek ugyanúgy hozzáférése van az intelligibilis valósághoz, mint a világléleknek, a különbség abban rejlik, hogy míg a világlélek fáradozás nélkül tudja irányítani a mindenség testét, az emberi lélekhez tartozó partikuláris test sérülékeny, a róla való gondoskodás kényszerei eltéríthetnek bennünket a kontemplációtól. A világlélek ily módon nem határozza meg az emberi lelket kauzálisan, csupán kijelöli számára azokat a kereteket, melyek között a kozmikus rendszerben kifejtheti autonóm működését.

A lélek testetlenségéből és az egyéni léleknek a világlélektől való függetlenségéből adódik, hogy az ember racionális lelke nincsen alávetve sem a fizikai világ kauzális rendjének, sem pedig e rend közvetlen princípiumának, a világléleknek. Úgy tűnik, *A végzetről* írott értekezésben Plótinosz megelégszik azzal, hogy az emberi lélek autonómiáját *a természeti okok* szövedékének kényszerítő hatalmával szemben biztosítja,<sup>38</sup> melyet a fátummal azonosít (III.1.10, 8–10; vö. 8, 10). A csillagok oksági befolyása a „külső”, fizikai okság speciális esete (III.1.9, 3 sk.). Az asztrális determinizmus széles körben elterjedt nézet a korban. Plótinosz részint azzal védi a racionális lélek autonómiáját, hogy a csillagok *jelzik* ugyan a jövőbeli eseményeket, de nem feltétlenül okozzák őket, részint azzal, hogy a csillagok kauzális befolyása korlátozott, és csupán testünket, illetve lelkünk alsóbb funkcióit érinti.<sup>39</sup>

A III.2–3 fejtegetése megmutatja, hogy az emberi lélek autonómiáját a gondviselés sem veszélyezteti. A *pronoia* az Értelemből, a második hiposztázisból származó *logosz* (racionális mintázat), mely a Lélek magasabb szintjén jelenik meg, majd másodlagos formában az alacsonyabb pszichikus szinteken, az élőlényekben bontakozik ki, ily módon közvetít az értelemmel felfogható valóság és a látható mindenség között (III.3.4, 8–13; 5, 16–20). A gondviselés második, alacsonyabb, a fizikai világban ható formáját Plótinosz – középső platonikus hagyó-

<sup>38</sup> Vö. IV.4.40, 23 sk. és 43, 18 sk.: a lélek megszabadulhat a fizikai világ varázslatától, mely fogva tartja alsó részét.

<sup>39</sup> Adamson (2008) finom elemzést ad az idevágó szövegekről, melyek közül a *Vajon hatnak-e a csillagok?* (II.3 [51]) a legfontosabb.

mányt követve – azonosítja a fátummal.<sup>40</sup> A *pronoia* megnevezés hagyományosan az oksági rend értelmes voltát, célszerűségét hangsúlyozza, amit Plótinosz a világ rendjének a hiposztatikus Értelemről való függésével magyaráz. A gondviselés működése az egyéni lelkek autonómiáját azért nem csorbíthatja, mert azok vonatkozásában nem az efficiens ok szerepét tölti be. Plótinosz ezen összefüggésben a *pronoianak* két aspektusát hangsúlyozza: a gondviselés egyfelől rendező elv, törvény, mely megszabja, hogy az autonóm ágensek milyen helyre kerülnek a mindenségben, másfelől normatív szabály, olyan mérce, melyen lemérhetjük tetteink racionalitását és morális minőségét.<sup>41</sup>

Kérdéseink persze ezután is maradnak. Az emberi lelkek nem a világlélektől függenek, nem is annak részei, hanem külön szálon függenek a magasabb princípiumoktól – tehát okságilag is függetlenek lehetnek tőle. Ez azonban nem változtat azon, hogy az egyéni lelkeknek megvan a maguk „létrehozója”, mely ráadásul ugyanaz a princípium, amelytől a világlélek is származik. Miért nem tekintjük tehát e forrást, az egyetemes Lelket a világfolyamat egyedüli meghatározójának? Nem mondhatjuk-e, hogy az egyéni lélek és a világlélek egyaránt ennek, s rajta keresztül a magasabb hiposztázisoknak az instrumentuma? Márpedig Plótinosz a sztoikusokat épp azért kritizálja, mert elméletükben – legalábbis szerinte – a kauzális felelősségnek önmagában kellene megalapoznia a morális felelősséget, s az emberek valójában csupán a fátum eszközei.<sup>42</sup> Továbbá felvethetjük a karakter eredetének problémáját. Ha tetteink egyedi karakterünkben (és a külső körülményekből) következnek, jellemünk viszont szintén további, külső okok terméke, miért nem hárul át a felelősség rólunk a jellemünket kialakító okokra?

E kérdésekre közös válasz adható. A *gondviselésről* írott értekezésben Plótinosz maga is felteszi a kérdést: honnan az ember egyéni jellege, morális minősége (*ho toioszde; to poion en toisz étheszin*)? Két lehetőséget vesz számításba: a jellem milyenségéért vagy az ember létrehozóját,<sup>43</sup> vagy pedig a létrejött dolgot – tehát magát a kérdéses személyt – terheli a felelősség (III.3.3–7). A fejezet hátralévő

<sup>40</sup> III.3.5, 15sk., vö. Apuleius: *Platón és tanítása* 1.12; Pszeudo-Plutarkhosz: *A végzetéről* 572F–573B.

<sup>41</sup> Az utóbbi értelemben mondhatjuk, hogy a rossz cselekedetek a „gondviselés ellenére” történnek, az erényes tettek pedig „gondviselés szerintiek” (*para pronoian; kata pronoian*). Plótinosz hangsúlyozza, hogy egyik esetben sem a gondviselés által (*hüpo pronoiasz*) mennek végbe tetteink, hanem mi magunk vagyunk azok ágensei (III.3.5, 47–54). A gondviselés rendező tevékenységéhez lásd III.2.13.

<sup>42</sup> „Mert a képzetek is az előzetes okokat követik, a késztetések pedig a képzeteket, s pusztán a név lesz csak, hogy valami „rajtunk múlik” (*teph' hémnin*); hiszen attól, hogy *mi* vagyunk, akik késztetéssel bírnak, semmivel sem inkább létezik ilyesmi; s olyan lesz a mi működésünk, mint a többi élőlényé és az oktalan gyermekeké, akik vak késztetéseik után mennek, vagy az örjögöke: hiszen ezek is bírnak késztetéssel; sőt – Zeuszra! – a tűznek is vannak késztetései, és mindannak, ami saját felépítésének szolgája, és mozgása azt követi.” III.1.7, 14–21.

<sup>43</sup> Létrehozó alatt itt nem az ember apját, hanem a magasabb metafizikai princípiumokat kell érteni.

részében amellelt érvel, hogy a létrehozó nem hibáztatható azért, mert gyengébb és a rossznak kitett lényt hozott létre. Az emanáció törvényeiből következik, hogy a létrehozott mindig hitványabb a létrehozónál, és a világ teljessége is megkívánja a különböző rangú élőlények létrejöttét. Marad tehát a másik lehetőség: ha az embernek saját magának kellett volna hozzájárulnia ahhoz, hogy jobbá legyen, akkor övé a felelősség (III.3.3, 12 sk). Az emberi ágens természetére, jelleme nem tekinthető külső okok okozatának:

Ha az ember egyszerű volna – „egyszerű” alatt azt értem, hogy ha pusztán az lenne, amivé tették, s eszerint gyakorolna és szenvedne hatásokat –, nos, ez esetben nem volna kárhoztatható, ahogyan más élőlények sem. Valójában azonban az a helyzet, hogy egyedül az embert gáncsolják, ha hitvány, s ez talán ésszerű is így. Hiszen nem pusztán az, amivé tették, hanem van benne egy másik, szabad princípium (*arkhén allén eleutheran*) (III.3.4, 1–7).

Az ember önmagáért való felelőssége abban gyökerezik, hogy bizonyos módon *önmagát alkotja meg*. A lélek emiatt nem egyszerűen a magasabb metafizikai princípiumok produktuma, és emiatt nem hárítható át a felelősség azokra a tényezőkre, melyeknek részük volt jellemünk formálásában.

A lélek többféle értelemben is önmagát alkotja meg. (1) Az emberben lévő „szabad princípium” az emberi lélekkel összetartozó értelmet (*nusz*) jelölheti, mely egyszersmind az Értelem birodalmának is teljes jogú polgára (vö. V.3.3, 21–26). Az isteni Értelem szerkezetét megvilágító *geneszisz*-történetekben a második hiposztázis egyrészt az Egy művének számít, másrészt önmagát alkotja meg, amennyiben az Egyből származó meghatározatlan emanátum „visszafordul” önmaga felé, és maga alakítja ki saját szerkezetét.<sup>44</sup> Az önkonstitúció az intelligibilis világ minden tagjának sajátja, így az ember racionális lelkének is metafizikai jellemzője.<sup>45</sup> (2) Az ember részese az egyetemes racionalitásnak, de hogy él-e ezzel a lehetőséggel, hogy érvényesül-e életében az egyetemes Értelem hatása, a benne lévő diszkurzív ész (*logosz*) állapotától függ, melyet Plótinosz az emberben lévő szabad, noétikus princípium elsődleges befogadójának, szubsztrátumának tekint (III.3.4, 32–34). A *logoszt* nem csupán az előbbi metafizikai értelemben tekinthetjük önkonstitutívnek:

<sup>44</sup> Az Értelem keletkezéstörténetének ehhez az aspektusához lásd Emilsson 2007. 70–77. A *Nusz* és az emberi lélek autonómiájának végső forrása az Egy autonómiája: az első princípium „önmaga oka”, önmagát hozza létre, és ebben az egyes entitások – az emberi lelkek is – részesednek (VI.8.13, 10–29).

<sup>45</sup> Proklosz ezt a plótinoszi gondolatot formalizálta az „önmagukat létrehozó entitások” (*authüposztata*) tanában (*A teológia elemei* 40–51).

S azt, hogy [a *logosz*] ilyen és ilyen (*to toionde*), az előző életre vezethetjük vissza, például a megelőző események nyomán elhalványulhat az őt megelőző racionális elvhez képest, de később ismét felragyog (III.3.4, 34–37).

Ebben a pontban a lélekvándorlás platóni mítoszai mellett Plótinosz a morális fejlődés és felelősség arisztotelészi elméletére támaszkodik. Arisztotelész azzal érvel a felelősség mellett, hogy karakterünket, lelki beállítottságunkat (*hexisz*) tetteink alakítják ki, vagyis nem természeti adottságok, hanem racionális ellenőrzésünk alatt állnak, mi magunk vagyunk érte felelősek (*Nikomakhoszi etika* III. 7). Plótinosz az emberi lélek állapotát meghatározó rész, a diszkurzív ész minőségét szintén azon az alapon utalja felelősségünk körébe, hogy az saját tetteink függvénye. (3) Az értelmes lélek önkonstitúciójának tétele nem csupán valamely, a lélek természetére vonatkozó metafizikai vagy a morális fejlődéssel kapcsolatos pszichológiai tény megállapítása, mely tetteinkért való felelősségünk biztosítékául szolgál, hanem egyben norma és felszólítás rejlik benne. Az ember feladata, hogy húzódjon vissza önmagába, távolítsa el, ami felesleges, egyengesse ki, ami egyenetlen, csiszolja simára, ami durva, miként egy szobrász, egészen addig, míg szépnek láthatja önmagát, s ezzel alkalmassá válik az isteni princípiumokhoz való felemelkedésre.<sup>46</sup> Az autonómia magasabb szintjét, a szabadság ideálját önmagunk átformálásával érhetjük el.

#### IV. MORÁLIS FELELŐSSÉG ÉS NORMATÍV AUTONÓMIA

A sztoikus fátumelmélet körüli vitákban az emberi autonómiára, mely a személyes felelősség alapjául szolgál, forrásaink jobbára az *eph' hémín* (*in nobis esse; situm esse in nobis*) kifejezéssel utalnak: lehetséges-e a vizsgált elmélet fényében, vagy sem, hogy eselekedeteink rajtunk állnak, tőlünk függenek, hatalmunkban vannak? A szabadság (*eleutheria*) terminológiáját a sztoikusok egy másik problémával kapcsolatban használják. A szabadság az ideális böles előjoga, egyedül ő szabad, ő független teljes mértékben mind a külső körülményektől, mind pedig a belső pszichológiai kényszerektől. Egy befolyásos nézet szerint a normatív szabadság e fogalma a sztoikus determinizmus körüli vitában soha nem játszott szerepet. Az *eph' hémín* és az *eleutheria* terminológiája csak császárkori szerzőknél kapcsolódik össze. Az összekapcsolás kétféle módon történik. A peripatetikus Alexandrosz, valamint bizonyos középső platonikusok és keresztény szerzők a meghatározatlan választás szabadságát teszik a morális felelősség alapjává, és erre kezdik alkalmazni mindkét terminust, ezzel teljesen új értelmet adva nekik. Epiktétosz az *eleutheria*, a normatív szabadság hagyományos sztoikus értékét állítja középpontba. Nála az *eph' hémín* fogalma kritériumként szolgál: igyeke-

<sup>46</sup> I.6.9, 7–25. Az öntranszformáció és a szabadság kérdéséhez lásd Remes 2007. 179–212.

zetünket kizárólag olyan dolgokra érdemes irányoznunk, melyek „a mi hatalmunkban vannak” – vagyis belső attitűdjeinkre, vélekedéseinkre és vágyainkra –, csak ezen a módon érhetjük el a szabadságot. A kauzális determinizmus elméleti problémája kiesik látóköréből.<sup>47</sup>

A *végzetről* írott értekezés utolsó fejezeteiben Plótinosz érvelési módja e fogalomtörténeti megfontolások fényében igen különös. A kiinduló probléma az volt, hogy miként egyeztethető össze a világ eseményeinek oksági meghatározottsága és az emberi autonómia. A felelősség alapjául szolgáló autonómiára Plótinosz körülírásokat használ, egy helyütt pedig az *eph' hémin* hagyományos terminusát.<sup>48</sup> Mikor saját elméletének kifejtéséhez fog hozzá, a morális felelősség kauzális megalapozása után a lélek normatív szabadságát kezdi ecsetelni:

Amikor [a lélek] a külső dolgokból kifolyólag másmilyenné válik, és mintegy vak késztetés révén mozog, azt kell mondanunk, hogy cselekvése nem önkéntes (*hekuszion*), sem pedig állapota nem az; s akkor is, amikor önmagából kifolyólag rosszabb, s nem minden tekintetben helyes és nem uralkodásra hivatott késztetések irányítják. Amikor viszont mozgását saját értelme (*logosz*) vezérli tisztán és hatásoktól mentesen, nos, egyedül e késztetésről kell azt mondanunk, hogy a mi hatalmunkban áll (*eph' hémin*) és önkéntes (III.1.9, 4–11).

Plótinosz érvelésének azt lehetne ellene vetni, hogy egyszerűen tárgyat vált. Az, hogy az egyéni lélek másra vissza nem vezethető kauzális forrásként lehet a morális felelősség hordozója,<sup>49</sup> értelmes tézisnek tűnik az adott összefüggésben. De mit keres itt az értékelő szempont, miért mondhatja Plótinosz, hogy csak a lélek normatív helyes késztetése múlik rajtunk (*eph' hémin*)? Talán a hitványakért nem vagyunk felelősek?

Úgy vélem, e vád kivédhető. A platóni-arisztotelészi hagyományban gyakran találkozunk azzal a gondolattal, hogy egyedül a *theória* szabad, ez független leginkább minden külső körülménytől, szemben a gyakorlati cselekvéssel, amely bizonyos fokig kiszolgáltatott a külső körülményeknek.<sup>50</sup> Plótinosz e gondolatot a jó és a hitvány tettekért viselt morális felelősség megalapozásába is bekapcsolja. Nyilván megfontolt nézetéről van szó, hiszen a morális felelősséget megalapozó autonómiát több helyen összeköti a normatív szabadsággal. A racionális egyéni lélek „nem jelentéktelen princípium”, amely önmagában, tes-

<sup>47</sup> E bekezdésben Bobzien (1998. 330–345) fogalomtörténeti rekonstrukcióját foglalom össze.

<sup>48</sup> III.13, 27: „a mi sajátosan emberi tevékenységünk”, *to hémeteron ergon*; 4, 21: „mi mi vagyunk”, *hémeisz hémeisz*; 7, 15 *to eph' hémin*.

<sup>49</sup> III.1.8, 4–8. A szöveget fentebb, a II. fejezetben idéztem.

<sup>50</sup> Arisztotelész: *Protreptikosz* fr. 12 Ross; Augustinus: *A Szentháromságról* 14.9.12; Iamblikhosz: *Protreptikosz* 9, 52,16–54,5 Pistelli; Alkinoosz: *Platón tanítása* 2; Plótinosz VI.8.5.

ten kívüli állapotában „legnagyobb mértékben a maga ura, szabad, és kívül áll a kozmikus oksági soron” (III.1.8, 7–10). Két hasonló megfogalmazást találunk *A gondviselésről* szóló műben:

[A]z emberek is princípiumok. Legalábbis a nemes dolgokra törekszenek saját természetüknél fogva, és e princípium szabad (*autexusziosz*) (III.2.10, 18 sk.).

Hiszen nem pusztán az, amivé tették, hanem van benne egy másik, szabad princípium (*arkhén allén eleutheran*) (III.3.4, 6 sk.).

Utóbbi idézetünk szerint az ember azért mentes a külső determinációtól („nem az, amivé tették”), mert hozzáférése van az egyetemes racionalitáshoz. Kauzális státusa – az, hogy princípiumnak, eredeti, külső tényezők által nem meghatározott kauzális forrásnak számít – összefügg azzal, hogy részese a normatív racionalitásnak. Természete szerint nem az az egyetlen lehetősége, hogy a test kényszereit és szükségleteit kövesse, hanem függetlenítheti magát tőlük, s „a nemes dolgokra törekedhet”.<sup>51</sup> Az ember e kauzális státusa biztosítja, hogy tettei nemcsak rajta keresztül valósulnak meg, hanem valóban ő tekinthető forrásuknak, így felelősséggel tartozik értük.

Térjünk végül vissza röviden a fogalomtörténeti összefüggésekre. Az, hogy a léleknek kizárólag az értelmet követő késztetései vannak „a mi hatalmunkban” (III.1.9, 9–11), annyit jelent, hogy csak ezek valósítják meg a normatív autonómiát (nem pedig, hogy a hitványakért ne volnánk felelősek). Más szóval, a frázis itt a „szabad” (*eleutherosz*) szinonimája. Plótinosz *Az Egy akaratáról és szabadságáról* (VI.8 [39]) írott munkában nem csupán rendszeresen szinonimaként használja a két kifejezést, hanem az *eph' hémin* fogalmának elemzése gyanánt e normatív szabadságfogalmat bontja ki, melyet nyilvánvalóan alapvetőnek tekint.<sup>52</sup> A fátumról és a gondviselésről írott Plótinosz-értekezések vizsgálata – szemben a szakasz elején ismertetett fogalomtörténeti konstrukcióval – azt mutatja, hogy e normatív szabadságfogalomnak szerepe lehet a morális felelősség megalapozásában egy determinisztikus metafizikai keretben.

<sup>51</sup> Ez persze nem jelenti azt, hogy a lélek diszpozíciója ne döntené el, miként választ ténylegesen.

<sup>52</sup> Ehhez az értekezéshez lásd Bene 2003.

## IRODALOM

- Adamson, Peter 2008. Plotinus on Astrology. *Oxford Studies in Ancient Philosophy*. 35. 265–291.
- Amand, David 1945/1973. *Fatalisme et liberté dans l'Antiquité grecque*. Amsterdam, Hakkert.
- Bene László 2003. Akarat és szabadság a sztoikus és a platonikus hagyományban: Epiktétosz és Plótinosz. *Világosság*. 44. 107–121.
- Bobzien, Suzanne 1998. *Determinism and Freedom in Stoic Philosophy*. Oxford, Clarendon Press.
- Boys-Stones, George 2007. 'Middle' Platonists on Fate and Human Autonomy. In Robert W. Sharples – Richard Sorabji (szerk.) *Greek and Roman Philosophy 100 BC – 200 AD*. Institute of Advanced Study, University of London. *Bulletin of the Institute of Classical Studies*, Supplement 94. 431–447.
- Brisson, Luc – Jérôme Laurent 2002. Traité 3 (III.1). Notice. In Plotin, *Traité 1–6*. Szerk. Luc Brisson – Jean-François Pradeau. Paris, Flammarion. 143–147.
- Bugár István – Lautner Péter (szerk.) 2006. *Sors és szabadság. Az emberi autonómia problémája az antik filozófiában a Kr. u. II. századig*. Budapest, Kairosz.
- Emilsson, Eyjólfur Kjalar 2007. *Plotinus on Intellect*. Oxford, Clarendon Press.
- Graeser, Andreas 1972. *Plotinus and the Stoics*. Leiden, Brill.
- Kendeffy Gábor (szerk.) 1998. *Antik szkepticizmus. Cicero- és Sextus Empiricus-szövegek*. Budapest, Atlantisz.
- Leroux, Georges 1996. Human Freedom in the Thought of Plotinus. In Lloyd P. Gerson (szerk.) *The Cambridge Companion to Plotinus*. Cambridge, Cambridge University Press. 292–314.
- Phillips, John 1995. Plotinus and Iamblichus on *Proairesis*. *Ancient Philosophy*. 15. 135–153
- Plotinus 1964–1982. *Opera* I–III. Oxford Classical Texts. Szerk. Paul Henry – Hans-Rudolf Schwyzer (*editio minor*). Oxford, Clarendon Press.
- Plótinosz 1986. *Az Egyről, a szellemről és a lélekről. Válogatott írások*. Ford. Horváth Judit – Perczel István. Budapest, Európa.
- Sharples, Robert W. 1994. Plato, Plotinus, and Evil. *Bulletin of the Institute of Classical Studies*. University of London. 39. 171–181.
- Sharples, Robert W. 2007. The Stoic Background to the Middle Platonist Discussion of Fate. In Mauro Bonazzi – Christoph Helmig (szerk.) *Platonic Stoicism – Stoic Platonism. The Dialogue between Platonism and Stoicism in Antiquity*. Leuven, Leuven University Press. 169–188.
- Sleeman, John Herbert – Gilbert Pollet 1980. *Lexicon Plotinianum*. Leiden–Leuven, Brill – Leuven University Press.
- Somos Róbert (szerk.) 2005. *Középső platonizmus*. Budapest, Osiris.
- Steiger Kornél (szerk.) 1983. *Sztoikus etikai antológia*. Budapest, Gondolat.
- Vidrányi Katalin (szerk.) 1994. *Az isteni és az emberi természetről*. Budapest, Atlantisz.
- White, Michael J. 1985. *Agency and Integrality: Philosophical Themes in the Ancient Discussions of Determinism and Responsibility*. Dordrecht–Boston–Lancaster–Tokyo, D. Reidel Publishing Company.



SIMON JÓZSEF

# Az intelligibilitás okságától az okság intelligibilitásáig A kauzalitás Duns Scotus filozófiájában

E konklúzióval és a következők némelyikével kapcsolatban az *aktuális létezésre* vonatkozó kijelentéseket is előadhatnék. [...] Azonban én jobban szeretek konklúziókat és premisszákat előadni a *possibilis létezésről*.

Duns Scotus: *De primo principio*, cap. III. concl. 1

## I. BEVEZETÉS

Általánosan elfogadott az a vélemény, hogy a klasszikus filozófiai tradíció elsősorban magyarázó elvként kezeli a kauzalitás problémakörét, valamint hogy a kauzalitás kérdését először Hume teszi szoros vizsgálat tárgyává (vö. Huoranszki 2001. 85–86). E magyarázatjelleg a klasszikus elméletben a világban előforduló dolgok lételméleti struktúrájának magyarázatát jelenti. Minthogy az individuális dolgok általános sajátosságait a klasszikus tradíció legkülönbözőbb formái sem képesek episztemológiai szempontok figyelembevételével tárgyalni, e sajátosságoknak a megismerhetőségükből és intelligibilitásukból következő fennállása lesz az, ami a kauzális magyarázat tárgyát fogja képezni. Tanulmányom kiindulópontját az a tézis jelenti, hogy az intelligibilitás ontológiai értelemben erős kauzális képességgel rendelkezik, amit paradigmatikusan fogalmaz meg a *Liber de causis* szövege. E máig tisztázatlan szerzőségű mű – amely köztudottan Proklosznak *A teológia elemei* című művéből vett részletek montázszerű újrafeldolgozása – arab filozófiai környezetben keletkezik valamikor a 10. században; ám 12. századi latin fordítása után meghatározó jelentősége lesz a 13. századi skolasztika számára. Ezután nagyon röviden azt kívánom bemutatni, hogy Aquinói Tamás hogyan hangolja át a létezők sajátosságait és a világ kauzális rendje közötti párhuzamosságot gondolkodásának arisztotelianus keretei között, és hogy hogyan tart ki mégis e párhuzamosság mellett. Tanulmányomat Duns Scotus filozófiájának néhány tézisével zárom. Scotus a metafizikai possibilitás nagy hatású elméletének kidolgozásával az intelligibilis létezés területéről számúzi a kauzális magyarázóelvet (Knuutila 1993. 139–149). Nála az intelligibilis sajátosságok és a világ kauzális rendje olyannyira radikálisan eltér egymástól, hogy

szinte kétségessé válik a kauzalitásról való lételméleti beszédmód lehetősége. Scotus azonban még nem Hume:<sup>1</sup> e redukált kauzalitás egyáltalában vett elgondolhatósága még nála is lételméleti teorémákat fog implikálni. Így érkezünk el az intelligibilitás erős értelemben vett kauzalitásának újplatonikus téziséből a természetes folyamatokban megvalósuló kauzalitás intelligibilitásának kérdéséig.

## II. A *LIBER DE CAUSIS* ÉS AZ ELSŐDLEGES OKOK KAUZALITÁSA

Minden elsőrendű oknak nagyobb a hatása okozatára, mint a másodlagos általános oknak.

Amikor a másodlagos általános ok saját hatóképességét tekintve távol marad a dologtól, az elsődleges általános ok nem távolítja el saját hatóképességét tőle.

Ez azért van így, mert az elsődleges univerzális ok már azelőtt hatással van a másodlagos ok okozatára, még mielőtt arra a másodlagos univerzális ok hatást fejtene ki.

Amikor a másodlagos ok – amely [az elsődlegesre] következik – hatást gyakorol az okozatra, hatása nem vonja ki magát a felette álló első ok hatása alól.

Amikor a másodlagos ok elkülönül az okozattól, amely rá következik, akkor a felette álló elsődleges ok nem különül el attól, mivel az annak okaként lép fel.

Ezt a létezés, az élő és az ember példáján keresztül mutatjuk meg.

Ez azért van így, mert a dolognak léteznie kell először, s azután kell élőnek, majd embernek lennie.

Az élő az ember közeli oka, a létező pedig távoli oka.

A létezés tehát erősebb értelemben oka az embernek, mint az élő, mert az az élő oka, ez utóbbi viszont az ember oka.

Hasonlóképpen, amikor azt állítod, hogy az ember oka a racionalitás, akkor is erősebb értelemben oka az embernek a létezés, mint a racionalitás, minthogy az nem más, mint az okának az oka (*Liber de causis* 1966. 90–91).

<sup>1</sup> Ezúttal nem tárgyalom a késő középkori szkepticizmus szerzőinek Hume-ot előlegező gondolatmeneteit. E szkeptikus irányzat legfontosabb képviselője Nicolaus de Ultricurria (1298/99–1369). Vö. Borbély 2008. 301–302.

A *Liber de causis* bevezető tézisei az elsődleges és másodlagos okok megkülönböztetését az okság általános jellemzőinek figyelembevételével írják le. A szöveg értelmezésekor különbséget kell tennünk egyfelől a konkrét elsődleges és másodlagos okok szembeállítására, másfelől az általában vett elsődleges és másodlagos okok szembeállítására között. Konkrétan három elsődleges ok létezik a három újplatonikus hiposztázisnak megfelelően, ám ezek kauzális képessége az egymáshoz rendeződő okok általánosabb koncepciói szerint gyakorol hatást.

Ezen általános koncepció értelmében az individuális dolgok sajátosságainak hierarchikus rendezettsége egyben e sajátosságok kauzális összefüggérendszeré is. Annak magyarázata, hogy egy individuum bizonyos sajátosságból következik az, hogy egy ennél általánosabb sajátossággal is rendelkezik, kauzális jellegű, amennyiben ezen általános sajátosság az oka a konkrétabb meglétének az individuumon. Az egyre általánosabb sajátosságok felfelé haladó sorában e kauzális képesség egyre növekvő mértékű, olyannyira, hogy egy általános és egy konkrét sajátosság között közvetítő kauzális sajátosságok akár el is maradhatnak: az általános sajátosság képes ezek nélkül is kiváltani a több lépcsővel alacsonyabb rendű és konkrétabb sajátosság meglétét; a *Liber de causis* e viszonyt is kauzális hatásgyakorlásként írja le.<sup>2</sup>

A 13. századi „virágzó” skolasztika számára a világban előforduló dolgok ontológiai felépítésének e kauzális magyarázata válik általánosan elfogadottá. A latin Nyugat filozófusai változó mértékű tudatossággal néztek szembe azokkal a problémákkal, melyekért a téves módon arisztotelészi eredetűnek tartott<sup>3</sup> *Liber de causis*ban szóhoz jutó arisztoteléus és újplatonikus vonások összefonódása volt felelős. A probléma nagyon tömören úgy foglalható össze, hogy az okság pontosan annak az Arisztotelésznek a filozófiájában játszik döntő szerepet, akinek kategóriatanát egyáltalán nem értelmezhetjük úgy, hogy a magasabb rendű kategóriális meghatározottságok *okai* lennének az alacsonyabb rendűeknek. A *Liber de causis* fenti példájánál maradva: az *élőlény* nem oka az *értelmes lénynek*, jóllehet kategóriális szempontból a két sajátosság univok módon rendeződik egymáshoz.

<sup>2</sup> Ezúttal lemondunk az e tézisekben implicit módon benne rejlő és az arab filozófiában széles körben elfogadott okkaszionalizmus-koncepció tárgyalásáról. Lásd ehhez Perler–Rudolph 2000.

<sup>3</sup> Aquinói egyike volt azoknak, akik rámutattak arra, hogy a szöveg milyen nagy mértékben újplatonikus, közelebbről prokloszi eredetű. Vö. Denis J. Brand előszavával, in *The Book of Causes* 1984. 2.

### III. AQUINÓI TAMÁS ÉS A PRINCÍPIUMOK KAUZALITÁSA

Aquinói Tamás filozófiája tipikusan tekinthető abból a szempontból, ahogy itt – az arisztotelianus skolasztika korában – a dolgok sajátosságainak rendje és a világ kauzális összefüggésrendszere egymásba fonódik. 1258–59 körül Aquinói Tamás Párizsban Boëthius *De trinitate* című művének kommentálásakor a következő gondolatokat fogalmazza meg a metafizika tárgymeghatározásával kapcsolatban:

Tudnunk kell, hogy minthogy bármely tudomány valamely alája vetett genust (*genus subjectum*) ismer meg, ezért meg kell hogy ismerje e genus princípiumait, mivel a tudomány csakis a princípiumok ismerete által éri el tökéletességét, ahogy az a Filozófus által a *Fizika* elején előadottakból világos.<sup>4</sup> Azonban a princípiumoknak két genusa van. Némelyek ugyanis önmagukban teljes természetek, mindazonáltal mások princípiumai is, ahogy az égi testek az alacsonyabban elhelyezkedő testek princípiumai, és ahogy az egyszerű testek az összetett testekéi. S ezért ezeket nemcsak annyiban veszik szemügyre a tudományokban, amennyiben princípiumok, hanem annyiban is, *amennyiben önmagukban vett létezők*. [...] Más princípiumok azonban nem önmagukban teljes természetek, hanem pusztán a természetek princípiumai, ahogyan az egység a számé, a pont a vonalé, vagy ahogy a forma és az anyag a fizikai testté, amiből kifolyólag az ilyesfajta princípiumokat kizárólag az a tudomány tárgyalja, amely a princípiumok által meghatározottakra vonatkozik (Thomas de Aquino 1965. 192–193; q. 5 a. 4 in corp. Kiemelés tőlem – S. J.).

Aquinói e szakaszban azon princípiumokat vizsgálja, amelyek a metafizika mint a minden dolgok princípiumait vizsgáló tudomány tárgyát képezik. A princípiumokat két részre osztja, aminek megfelelően a metafizika tudománya is két részre oszlik: adottak ugyanis olyan princípiumok, amelyek azon túl, hogy ezeket egy adott tudomány bizonyítatlan előfeltételeiként felvesszük, önálló egzisztenciával is rendelkeznek, illetve adottak olyan princípiumok, amelyek nem rendelkeznek ilyen önálló egzisztenciával. A metafizika feladata nem pusztán abban áll, hogy princípiumokat vegyen fel a tudományok egyáltalában vett előfeltételeiként, hanem ezeket *önmagukban vett létezőkként* is meg kell vizsgálnia. E vizsgálat érdekében a metafizikán belül fel kell vennünk egy olyan külön részt, amely e kérdést tisztázza. A metafizika ilyen felosztását meg kell

<sup>4</sup> Arisztotelész: *Fizika* 184a12–14. A *Fizikának* a szóban forgó hely kontextusát képező bevezető passzusa (184a10–b13) az egyik legfontosabb tudomány módszertani hely a *Corpus Aristotelicum*-ban, különösen a reneszánsz arisztotelianusok páduai tradíciója hivatkozik rá mint a tudományok tapasztalati kiindulópontját igazoló szövegrészletre.

engednünk, hiszen a természettudomány különböző diszciplínái esetében is megfigyelhetjük ezt a felosztást, amennyiben a felsőbb régiókban elhelyezkedő testek kutatása azon kívül, hogy tisztázza a Hold alatti szféra testeire irányuló kutatás princípiumait, e princípiumok létezését is tematizálja. Másképp fogalmazva: a Hold alatti szféra testi szubsztanciáinak princípiumaira irányuló tudomány nemcsak azt tárgyalja, hogy mi az összetett testek princípiuma, hanem azt is, hogy e princípiumok kifejezetten azonosak a Hold feletti szféra égitesteivel. Bár Aquinói itt még nem mondja ki, gondolatmenete mégis abba az irányba halad, hogy a metafizikának egyáltalában nem volna tárgya, ha a létezés általa vizsgált princípiumainak nem volna valamiféle önálló egzisztenciája.

Ahogy bármely egyes meghatározott genusra nézve adottak bizonyos közös princípiumok, amelyek ezen genus összes princípiumaira kiterjednek, úgy az összes létező is – amennyiben részesülnek a létezőben – rendelkezik bizonyos princípiumokkal, amelyek minden létező princípiumai. Avicenna véleménye szerint a *Sufficienciában* ezen princípiumokról kétféleképpen állíthatjuk, hogy közösek [Avicenna 1992. 22–23; I c. 2 f. 14va 6–30]. Egyik módon a *predikáció* szerint, ahogyan a következőt állítom: a forma közös az összes formára nézve, mivel bármelyikről predikálható. Másik módon a *kausalitás* szerint, ily módon állítjuk azt, hogy a szám szerint egy Nap minden keletkezhetőre vonatkoztatva princípium. Az összes létező ugyanis nem csak az első mód szerint rendelkezik közös princípiumokkal, amire a Filozófus a *Metafizika* XI. [Kappa] könyvében felhívja a figyelmet,<sup>5</sup> tudniillik, hogy minden létező ugyanazon princípiumokkal rendelkezik az *analógia* módján, hanem a második mód szerint is, oly módon, hogy mindennek olyan dolgok legyenek a princípiumai, amelyek szám szerint egyek. Ily módon az akcidensek princípiumai visszavezetnek a szubsztancia princípiumaira, és a pusztulható szubsztanciák princípiumai visszavezetnek a pusztulhatatlan szubsztanciákéra, és így egy bizonyos fokozatosság és rend szerint minden létező visszavezet bizonyos princípiumokra. És mivel annak, ami a létezőség princípiuma minden dologban, a legmagasabb rendű létezőnek kell lennie, ahogy azt a *Metafizika* II. [alfa] könyve állítja [1049b4–1051a33], ennél fogva az ilyesfajta princípiumoknak a legteljesebbeknek kell lenniük, *emiat aztán a legmagasabb rendű aktualitásban kell lenniük*, hogy semmiféle vagy csak a legesekélyebb mértékű potencialitással bírjanak, mivel az aktualitás magasabb rendű a potencialitásnál, ahogy azt a *Metafizika* IX. [Théta] könyve mondja [993b26–31]. Ennél fogva teljes mértékben mentesnek kell lenniük az anyagtól és a mozgástól, amely utóbbi a potencialításban egzisztáló

<sup>5</sup> Helyesen: XII. [Lambda] könyvében, vö. Arisztotelész: *Metafizika* 1070a31–33 és 1071a30–35.

aktualitása. És ilyenek az isteni dolgok, *mivel ha az isteni valahol egzisztál, akkor egy ilyen természetben* – azaz immateriális és mozgatatlan természetben és a legmagasabb rendű módon – *egzisztál*, miként azt a *Metafizika* VI. [Épsilon] könyve állítja [1026a20] (Thomas de Aquino 1965. 193–194. Kiemelések tőlem – S. J.).

Világos, hogy ahogy „*bármely egyes genusra nézve adottak bizonyos közös princípiumok*”, úgy adottak olyan princípiumok is, „*amelyek minden létező princípiumai*”. Aquinói ismét megkülönböztetést vezet be a metafizika tárgyát illetően: ezen minden létezőre közös sajátosságok egyfelől a *predikáció*, másfelől a *kauzalitás* szempontjából közősek. Bármennyire is Avicennára támaszkodva vezet be e megkülönböztetést, a distinkció éppen az arab filozófus metafizika-felfogásával ellentétes irányú. Tamás ugyanis a hangsúlyt arra helyezi, hogy azon túl, hogy az arisztotelészi analógia elve alapján a létezés általános princípiumait minden egyes létezőről predikálhatjuk – hiszen a világ dolgai a kategoriális rendben adottak –, fel kell vennünk e princípiumok egzisztens létezését is, „*oly módon, hogy mindennek olyan dolgok legyenek a princípiumai, amelyek szám szerint egyként léteznek*”. A princípiumok létezése teljes szubsztanciális létezés. A metafizika fő része így épphogy nem a pusztán – azaz *predikatív* „létmóddal” – princípiumokként felvetek tematizálása, hanem azok dologszerű létezésének tisztázása. Aquinói az (onto)teológiaként felfogott metafizika tárgymeghatározásának arisztotelészi kulshelyeire hivatkozva a metafizikai princípiumok kifejezett egzisztens-szubsztanciális létmódja mellett foglal állást. A metafizika ilyen tárgymeghatározása kifejezetten a létezés általános princípiumainak *kauzalitás* szerinti közösségén alapul: a metafizika primér szubjektuma isten mint legmagasabb rendű *ok*. A predikáció szerinti közösség – mint a létező mint létező transzkategoriális sajátosságainak általánossága – kizárólag az isteni általános kauzalitás által léphet fel a világ szubsztanciáiban.

Tehát az ilyesfajta isteni dolgokat, mivel az összes létező princípiumai és nem kevésbé önmagukban teljes természetek, kétféleképpen lehet tárgyalni: az egyik módon annak megfelelően, ahogy az összes létező princípiumai, a másik módon annak megfelelően, ahogy önmagukban bizonyos dolgok. Mivel azonban az ilyen első princípiumok bárhogy is léteznek, *önmagukban a leginkább ismertek*,<sup>6</sup> és értelmünk mégis úgy vonatkozik rájuk, mint ahogyan az éjszakai sötétséghez szokott szem a nap fényéhez, miként azt a *Metafizika* II. [alfa] könyve állítja [993b9–11], ennél fogva a természetes ész fénye által kizárólag oly módon jutunk hozzájuk, hogy az okozatokból vezetettünk ismeretükhöz. És ez az a mód, ahogy a filozófusok hozzájuk jutnak, ami világos a Róm. 1-ből: *a mi istenben láthatatlan [...] az ő*

<sup>6</sup> Vö. *Második Analitika* 71b9–72b4, *Fizika* 184a10–b14.

*alkotásaiból megértetvén megláttatik.*<sup>7</sup> Ebből következően az ilyen isteni dolgokat a filozófusok kizárólag mint az összes dolgok princípiumát tárgyalják. Ennélfogva ezek abban a tanban kerülnek tárgyalásra, amelyben ezek az összes dolgok princípiumaiként lesznek tételezve, és amelynek szubjektuma a *létező mint létező*: és ezt nevezik ők [a filozófusok] isteni tudománynak. Van ugyanis egy másik módja is az efajta dolgok megismerésének, nem aszerint, ami az okozatokból nyilvánvaló, hanem aszerint, ahogyan ezek önmagukban nyilvánvalóak. Ezen módon állította az apostol az 1Kor 2[, 11–12]-ben: *Azonképpen az isten dolgait sem ismeri senki, hanemha az istennek Lelke. Mi pedig nem e világnak lelkét vettük, hanem az istenből való Lelket; hogy megismerjük azokat, a miket isten ajándékozott nekünk. Ugyanott: Nekünk azonban az isten kijelentette az ő Lelke által.* [1Kor 2,10.] Ezen a módon az isteni dolgok aszerint kerülnek tárgyalásra, ahogyan önmagukban szubjektálnak, és nem pusztán aszerint, ahogyan a dolgok princípiumai (Thomas de Aquino 1965, 194–195. Kiemelések tőlem – S. J.).

Jóllehet e princípiumok „*önmagukban a leginkább ismertek*” (az arisztotelészi *té [...] phüszei gnórimóteron a Második analitika 71b35–72a1* alapján), ismeretelméleti szempontból mégsem vagyunk képesek számot adni arról, hogy hogyan jutunk e princípiumok önmaguk általi ismeretéhez. Aquinói Arisztotelész Platón-allúziójával megismerőképességünket az éjszakai sötétséghez szokott szemhez hasonlítja, amely a nappali fényben képtelen meglátni a dolgokat. Ismereteinket „a természetes ész fénye által” a princípiumokkal kapcsolatban is az okozatok éjszakai sötétjéből kiindulva „az ő alkotásaiból megértetvén” nyerjük. Csak „istennek Lelke” képes a princípiumoknak nem okozataikon nyugvó, hanem önmaguk általi ismeretére. Az emberi megismerőképesség számára pontosan a kauzalitás elvének messzemenő metafizikai alkalmazása által nyert princípiumok biztosítják azt, hogy a *létező mint létező* a metafizika szubjektuma legyen. Aquinói azonban rögtön hozzáfűzi: „és ezt nevezik ők [a filozófusok] isteni tudománynak”. Amikor azt írja, hogy „az ilyen isteni dolgokat a filozófusok kizárólag mint az összes dolgok princípiumát tárgyalják”, akkor egyfelől ugyan arra utal, hogy a filozófusok itt zárójelbe teszik a teológiai istenfogalmat, másfelől azonban félreérthetetlenül a tudunkra adja, hogy a létező mint létező csakis istennek a világra vonatkozó *kauzális viszonya* által lehet a metafizika szubjektuma. Nem lehetséges általános ontológia a racionális-filozófiai teológia előzetes tematizálása nélkül. Úgy tűnik, hogy az általános ontológia a racionális-filozófiai teológia foglya marad mindaddig, amíg a dolgok létjellemzőiket egy legfőbb ok kauzális hatásgyakorlásának köszönhetik.

<sup>7</sup> Róm 1,20. A bibliai idézetek a Károlyi-fordítás alapján.

## IV. DUNS SCOTUS

## 1. Kauzális és metafizikai lehetőség

E konklúzióval és a következők némelyikével kapcsolatban az *aktuális létezésre* vonatkozó kijelentéseket is előadhatnák a következőképpen: Valamely természet hatóok, mivel valamely természet hatóokság okozata, mivel valamely természet elkezd létezni, mivel valamely természet a mozgás végpontja és kontingens. Azonban én jobban szeretek konklúziókat és premisszákat előadni a *posszibilis létezésről*. Ugyanis feltéve, hogy az aktuális-sal kapcsolatos érvek helytállóak, a posszibilisre vonatkozóak is helytállóak lesznek, ám ez megfordítva nem áll fenn. Továbbá, az aktuálisra vonatkozó érvek, bár eléggé nyilvánvalóak, mégis *kontingensek*, míg a posszibilisre vonatkozóak *szükségszerűek*. Azok az egzisztens létezőre vonatkoznak, ezek sajátosan a quiditativ értelemben felfogott létezőre tartozhatnak (Duns Scotus 1639. 229. *De primo principio*, cap. III. concl. 1.).

Amikor Duns Scotus 1305 körül ezen sorokat papírra veti *De primo principio* című műve harmadik fejezetében, akkor egy olyan preferenciát tár fel, mely nemcsak a művében megfogalmazott érveit, hanem teljes filozófiai stílusát és attitűdjét jellemzi. A személyes hangvételű állásfoglalás arra vonatkozóan, hogy „én jobban szeretek konklúziókat és premisszákat előadni a *posszibilis létezésről*” – értsd: az *aktuális létezésre* vonatkozó érvekkel szemben –, félreérthetetlenül adja az olvasó értésére, hogy a scotusi filozófia fő témája és egyben legfőbb filozófiatörténeti nívója a modalitás problémakörének újragondolása. A modális létezés jellemzőinek explikálása jól láthatóan éppen az aktuális létezés területén és az itt megnyilvánuló kauzális viszonyokat illetően fog nehézségeket „okozni”. E zavart jelzi Scotus fenti megjegyzése is: A *De primo principio* gondolatmenetének – mely a középkori filozófia talán legrészletesebben kifejtett érvelése isten létezésével – vezérfonala a hatóoki kauzalitás struktúráján keresztül vezet isten létezésének belátásához. Idézetünk bevezető soraiban azonban Scotus mintha épp egy ilyen érvelés trivialisáról beszélne: a kizárólag az egzisztens-aktuális létszféra figyelembevételével megfogalmazott érv helyett a posszibilis létezéssel kapcsolatos argumentációt javasol.

Scotus – az európai filozófia történetének nagy szerzőivel összhangban – azon nyilvánvaló tény magyarázatát keresi, hogy a világban individuális dolgok generalizálható sajátosságokkal rendelkezve léteznek. E magyarázatnak – abban a formában, ahogyan Arisztotelész a problémát felvetette – egyben az általa első filozófiának nevezett metafizika első tárgyáról kell szólnia. Az antik filozófus két alapvető magyarázattípust hagyott örökül a metafizika tárgyát illetően, és bölcs hallgatásba burkolózott az ezek között fennálló prioritást illetően. Az egyik megoldás szerint kijelölhető egy meghatározott létező, amely individuális egzisztens-



ciája szerint mint a létezés *maximális mozzanata* magába foglalja minden dolgok lételvét. Úgy tűnik, hogy Arisztotelész emellett száll síkra, amennyiben a metafizika feladatát úgy állítja be, mint az *okok* nemeinek vizsgálatát, amely végül is ahhoz a belátáshoz vezet, hogy a metafizika tárgya isten, azaz a legmagasabb rendű ok (*Met.* VI [Epsilon] 1, 1025b1–1026a22). A metafizika e meghatározás szerint (racionális) (onto)teológia (vö. *Met.* I [Alfa] 982a1–3, 982b9–10, 983a8–10; *Met.* IV [Gamma] 1003a26–32; *Met.* VI [Epsilon] 1025b3–4, 1026a18–23, 1028a3–4; *Met.* XI [Kappa] 1064b1–6). Ezzel szemben több helyen is azt olvassuk a *Metafizikában*, hogy a metafizika tárgya a létező mint létező, illetve azon sajátosságok, amelyek a létezőhöz mint létezőhöz járulnak hozzá. E meghatározás szerint a létező mint létező minden általánosan vett dologin fellelhető *minimális mozzanata* jelöli ki a metafizika tárgyterületét, a metafizika így mindenekelőtt *ontológia* (*Met.* IV [Gamma] 1003a21–25; *Met.* VI [Epsilon] 1028a21–24; *Met.* XI [Kappa] 1060b31–1061b17).

Arisztotelész első megoldási javaslata, miszerint a metafizika univerzális diszciplínájának tárgya mindenekelőtt egy kitüntetett létezőben jelölhető meg, egészen a 13. század végéig uralkodó felfogás marad. E kitüntetett létező isten, a metafizika tudománya így mindenekelőtt racionális vagy természetes teológia, (onto)teológia. Annak ellenére, hogy isten természetesen az összes kifejtett rendszerben meghaladja a világ kategoriális rendjét (szolgáljon annak alapjául akár egy arisztoteléanus természetfilozófia, akár az ideák és a részesezés platonikus fogalmai által meghatározott kettős ontológiai rend, vagy akár egy radikálisan isten akarától függő augusztiniánus-keresztény létrend), mégis mint individuum bírja a világ lételvét. E felfogás szerint isten a világ individuális dolgainak létjellemezőit a világra vonatkozó *kauzális* viszonya alapján „osztja le”. E felfogást szemléltettük fent a *Liber de causis*ből és az Aquinói Tamástól származó részletekkel.

A másik javaslatot, amely a metafizika tárgyaként a létezőt mint létezőt állítja be, a hellénisztikus és későbbi tradíció nagy része figyelmen kívül hagyja. E negligálás elméleti motívumait helyhiány miatt most nem ismertette<sup>8</sup> arra kell utalnunk, hogy Duns Scotus a hellénizmuskori alexandriai Ammóniosz-iskola és Avicenna kísérletei után először veti fel a metafizika tárgymeghatározásával kapcsolatos „másik” arisztotelészi megoldást, miszerint a metafizika mindenekelőtt a létezőről mint létezőről szóló tudomány. Szintén figyelmen kívül hagyva a létező mint létező transzkategoriális voltából fakadó klasszikus apóriákra adott választ,<sup>9</sup> röviden leszögezhetjük, hogy Scotus szerint az *ens inquantum ens* a létezés *univok* sajátossága, amely áthatja a teljes kategoriális rendet.<sup>10</sup> Annak tárgyalásáról is lemondva, hogy milyen episztemológiai szempontok szólnak a léte-

<sup>8</sup> Lásd részletesen: Zimmermann 1965.

<sup>9</sup> Duns Scotus Ed. Vat. III. 81–85 (Ord. I d. 3 p. 1 q. 3 nn129–136).

<sup>10</sup> Duns Scotus Ed. Vat. IV. 205–206 (Ord. I d. 8 p. 1 q. 3).

ző mint létező és az alatta kibomló kategorialitás formális ontológiája mellett,<sup>11</sup> rögzíthető, hogy Scotus szemében a kategorialis rend a világra nem egy legfőbb ok kauzalitása folytán jellemző, hanem azért, mert az *ens inquantum ens* mint intencionális objektum explikálódik a világban,<sup>12</sup> méghozzá oly módon, hogy ezen explikáció teljesen indifferens a világ kauzális viszonyaira nézvést.

A létező mint létező esetében episztemológiai és ontológiai szempontok egymásba játszásáról beszélhetünk. *Episztemológiai* szempontból a létező mint létező minden megismerés egyáltalában vett tere – (nem is olyan) halvány előképe ez az ismeretek lehetőségfeltételére apelláló újkori transzcendentálfilozófiai *Bedingung*nak (vö. Honnfelder 1990) –, és ily módon a világ dolgainak leggyengébb *ontológiáját* sajátossága, mely ugyanakkor a legáltalánosabb: minden individuális dologon fennáll az *ens inquantum ens* minimális mozzanata. A két aspektus abban a filozofémában ér össze, miszerint a létező mint létező egy olyan transzkategorialis esszenciális diszpozíció, mely az ellentmondásmentesség által artikulálódik.<sup>13</sup> Az episztemológiai szempontból ellentmondásmentes bejelentheti igényét bizonyos ontológiai tételezettségre. Ha ezt a kis engedményt megteesszük, akkor az ontológia nem lesz többé az isten létezését tematizáló racionális teológia függvénye, hanem az általános létjellemezők egy önálló általános ontológia keretein belül tárgyalhatóak; és ezen létjellemezők istenre is érvényesek lesznek.

Közelebb lépve témánkhoz: a létezőnek mint létezőnek az ellentmondás-mentesség által tételezett esszenciális diszpozíciója tulajdonképpen nem más, mint az egyáltalában vett *lehetőség*. Scotus Arisztotelész *Metafizikájának* a potencialitás és aktualitás problémáját tárgyaló IX. [I'héta] könyvéhez írt kommentárjában fejti ki sajátos koncepcióját a *metafizikai lehetőség*ről.

A potencia egyfelől a létező bizonyos létmódját jelenti. Másfelől speciálisan a princípium elvét vezeti be. Kétséges, hogy e kettő közül melyik számára lett bevezetve előbb a név, és melyikből vittük át a másikra. Ha mégis előbb a létező bizonyos módjának megjelölésére alkalmazzuk, mivel az kizárólag a létezését lehetővé tevő princípium által járulhat egy ilyen létezőhöz, ennek megfelelően a potencia nevét átvihetjük a princípiumra, mintegy arra, ami által a possibilis képes létezni, nem mint „ami által” formálisan, hanem kauzálisan (*non 'quo' formaliter, sed causaliter*). Hasonlóan, ha előbb azon princípiumra alkalmazzuk, amely által a dolog létezhet, akkor át lehet vinni a létezés hozzá hasonló módjának – amelyet az okozat az okban bír – általános megjelölésére (Duns Scotus: Op. Phil. IV. 512. *In Metaph.* IX, q. 1–2 n14).

<sup>11</sup> Duns Scotus Ed. Vat. XVI. 250 (Lect. I d. 3 p. 1 q. 1–2 nn68–69).

<sup>12</sup> Az intencionalitás problémájához Duns Scotusnál lásd Perler 2002. 217–230.

<sup>13</sup> Klasszikus hely: Duns Scotus Ed. Viv. XXV. 113–115 (*Quodl.* q. 3 n2).

A potencia kifejezés jelentésátviteli alternatíváinál fontosabb a két elv pontos megkülönböztetése. A potencia filozófiai terminusa egyrészt „a létező bizonyos módját”, másrészt „speciálisan a princípium elvét” jelenti. A kölcsönös jelentésátvitel lehetőségét tárgyalva Scotus pontosítja a két potencia különbségét: A potencia mint „a létező bizonyos módja” egy dolgot úgy jellemez, mint „ami által» [az] formálisan” lehetséges, míg a potencia mint „speciálisan a princípium elve” egy dolgot úgy jellemez, mint *kauzális* lehetőséget, amely a világban lefutó ok-okozati sorokba illeszkedik. A princípium e részletben kifejezetten kauzális princípiumot jelent, azt a lehetőség szerinti létezést, melyet „az okozat az okban bír”. A lehetőség fogalma nem merül ki a potencia kauzális értelmében. Pontosan a nem kauzális jellegű lehetőség lesz a metafizikai kutatás tárgya:

Itt tehát a közősebb értelemben vett potenciát kell szemügyre vennünk, tudniillik aszerint, ahogy az a létező bizonyos önmagában vett és a princípium elve nélküli modusát jelenti. És mivel a metafizikát művelő kutató a létezőt és annak sajátosságait, ebből következően az ily módon felfogott potencia a metafizikát művelő kutatására tartozik. Ennélfogva az előadás rövidségére tekintettel ezt *metafizikai potenciának* lehet nevezni (Duns Scotus: Op. Phil. IV. 512–513. uo. n16).

A *metafizikai lehetőség* tehát a létező önmagában vett és a princípium elve nélküli modusa, ahogy egy dolog önmaga formális-esszenciális konstitúciója folytán lehetséges. A metafizikai lehetőség többféleképpen artikulálható. Egyfelől a létező mint létező közvetlen sajátossága, ahogy Scotus fogalmaz: „a *possibilis* felecsérelhető a teljes létezővel, mivel semmi sem létező, aminek elve ellentmondást foglal magában”, másfelől azt nevezzük „*lehetséges létező*nek, ami pusztulható entitással rendelkezik”.<sup>14</sup>

A kontingencia a téridőbeli világban létező dolgok metafizikai sajátossága. Már itt is megfigyelhető a kauzalitás formális megközelítése, amennyiben Scotus a kauzális lehetőséget a kontingencia formális sajátosságaként is jellemzi. E formális nézőpont szerint nem azért kontingens valamely dolog, mert képes egy kauzális sor valamely láncszemének funkcióját betölteni, hanem azért képes egy kauzális sor valamely láncszemének funkcióját betölteni, mert már eleve formálisan kontingens.<sup>15</sup> A kauzális funkció abból következik, hogy minden olyan időpillanatban, amikor valamely kauzális viszony megvalósul, a világ tényei lehetnének másképp is, mint ahogyan aktuálisan fennállnak.

Itt érkezünk el tulajdonképpeni problémánkhoz. Ha Scotus ily mértékben kifejtethetőnek tartja a *potencia* nem kauzális, hanem formális-metafizikai értelmét, akkor hogyan fogja értelmezni a kauzalitás kérdését? Úgy tűnik, az a tény,

<sup>14</sup> Duns Scotus: Op. Phil. IV, 515 (*In Metaph.* IX, q. 1-2, n21).

<sup>15</sup> Vö. *Lectura* I d. 39, kommentált, kétnyelvű kiadása: Duns Scotus 1994.

hogy a világ individuális dolgai bizonyos generalizálható sajátosságokkal rendelkeznek, nagyon jól magyarázható azon formális létezés által, amelyet e sajátosságok elgondolhatósága implikál. Természetesen annak magyarázata, hogy egy fehér asztal miért fehér, és miért nem barna – vagyis miért nem olyan színnel rendelkezik, amelynek fennállása az aktuálisan fennálló fehérrel szinkron módon lehetséges –, nem nélkülözheti azon aktuális kauzális viszony felmutatását, amelynek eredményeképp az asztal fehér színű lett. Ám a tradicionális felfogás szerint csak akkor rendelkezik az aktuálisan fehér asztal a barnává válás lehetőségével, amikor ott állok mellette barna festékesdobozzal és ecsettel. Scotus szerint azonban a fehér asztal akkor is rendelkezik a barnává válás lehetőségével, ha a világ kauzális viszonyai közepette ez az asztal sohasem lesz barna. Ellentmondásmentesen el tudom gondolni ezt a fehér asztalt barnaként, ami azt jelenti, hogy a létező mint létező alatt kibomló kategoriális rend megengedi ezen asztal barnaságának lehetőségét. A kauzalitásra vonatkozó kérdést úgy tehetjük fel, hogy vajon a kauzalitás mit képes ezen formális kontingenciával szemben teljesíteni. Mit mondhatunk a kauzalitásról akkor, ha úgy tűnik, hogy Scotus kitart amellett, hogy a legdetermináltabb kauzális viszony okozata is messzemenően kontingens marad, mégpedig amiatt, hogy ellentmondásmentesen elgondolható egy tőle különböző megvalósulatlan lehetőség?

## 2. A kauzalitás törvényei Duns Scotusnál

A kauzalitás kérdésköre elsősorban az istenérvek négy, egymással nagy mértékben egyező megfogalmazásában játszik fontos szerepet – vizsgáljuk meg akár az *Ordinatio* (I d. 2 p. 1 q. 1–2.), akár a *Reportationes* (I d. 2 q. 1–4), akár a *Lectura*<sup>16</sup> (I d. 2 p. 1 q. 1–2) megfelelő fejezetét, vagy akár a teljes *Tractatus de primo principio* érvelését.<sup>17</sup> Scotus istent előzetesen mint *végtelen létezőt* definiálja (mert istent az *ens inquantum ens* végtelen modális transzcendententáléja alatt gondoljuk el), s e definíció mellett még bizonyításra szorul, hogy valóban létezik-e egy ilyen sajátossággal rendelkező létező.<sup>18</sup> A bizonyítás *a posteriori* jellegű és a

<sup>16</sup> *Lect.* I d. 2 p. 1 q. 1–2.

<sup>17</sup> Duns Scotus 1639. 210–259. A *De primo principiónak* nincs mértékadó modern kritikái kiadása, ami nem utolsósorban annak köszönhető, hogy ma is hitelesnek tűnő kéziratokban maradt ránk – ami egyáltalán nem jellemző Scotus egyéb munkáira –, így az egyébként máshol megbízhatatlan 17. századi Wadding-kiadás ez esetben használható. Vö. Vos 2006. 141–142. A szöveg kiváló német és angol fordításokban is hozzáférhető: Duns Scotus 1974; Duns Scotus 1966 (több kiadás).

<sup>18</sup> Nagyon fontos e ponton megjegyezni, hogy Scotus istendefiníciója (*ens infinitum*) nem *a priori* jellegű, ahogy azt Anzelmnél látjuk. A végtelenség úgynevezett diszjunktív transzcendententáléja a létező mint létező intencionált esszenciális diszpozíciója alatt explikálódik. Scotus az intencionális kategorialitás és transzkategorialitás intelligibilis tartalmait *a posteriori* tartalmaknak tekinti, annak ellenére, hogy isten sohasem közvetlen tárgya az emberi megismerőképességnek. Scotus a következő „színezéssel” (*coloratio*) tartja elfo-

végtelen létezőnek a teremtett dolgokra nézvést relatív sajátosságain alapul, melyek két típusba sorolhatóak. Egyfelől kauzális sajátosságokról (*prorietates causalitatis*), másfelől eminens sajátosságokról (*prorietates eminentiae*) beszélhetünk. A teremtett dolgok között tehát alapvetően két olyan rend van, melyek valamiféleképpen vonatkozásban állnak a mérhetetlenül transzcendens isten bizonyos sajátosságaival: a kauzalitás és az eminencia rendje. Az eminencia, a dolgok sajátosságait jellemző tökéletesség rendje, megegyezik a világ kategoriális rendjével. Az eminencia rendje a sajátosságok intelligibilis lehetőségének rendje, mely – mint fentebb láttuk – indifferens az egzisztens kauzalitás viszonylatában.

Scotus a következő kauzalitástörvényekkel dolgozik:

n49 [A] *per se* rendezett okok között a későbbi ok, amennyiben további okozatot vált ki, függ az őt megelőzőtől. A *per accidens* rendezett ok nem függ ilyentől [amikor okoz], jóllehet létezésében vagy valami mást tekintve függhet.

n50 [A] *per se* rendezett okok kauzalitása más elvből és más rendből származik, mivel magasabb rendű és tökéletesebb; az akcidentiálisan rendezettek esetében azonban nem. E különbség az elsőből következik, mivel egy ok sem függ a kauzális hatásgyakorlásban olyan [megelőző] októl, melynek elve megegyezik az övével, mert valami kauzális kiváltásához elegendő egyetlen elv.

n51 [A] *per se* és esszenciálisan rendezett okok mindegyike egyszerre szükségzerű követelmény az okozáshoz, máskülönben valamely esszenciális kauzalitás hiányozna az okozat számára; az akcidentiálisan rendezettek esetében ez nem így van, mert a kauzális hatás kiváltása nem követeli meg ezt a szimultaneitást (Duns Scotus: Ed. Vat. II. 154–155. *Ord.* I d. 2 p. 1 q. 1–2 nn49–51).

Az n49 az esszenciális kauzalitás láncolata tagjainak *erős függését* mondja ki, az n50 az egymáshoz rendeződő esszenciális hatóokok *kategoriális eltérését*, míg az n51 azt a tézist hangoztatja, hogy egy esszenciális kauzális viszony elengedhetetlen feltétele az ily módon egymáshoz rendeződő okok *szimultán hatáskifejtése*. E kauzalitástörvények megfogalmazása egy *a posteriori* istenérv keretein belül történik meg: Scotus általuk kívánja kiküszöbölni a hatóokok sorában fellépő végtelen regresszust. Mindhárom tézis megkülönbözteti az okok és okozatok *per se* (esszenciálisan) és akcidentiálisan rendezett sorát. Scotus intenciója szerint az okok esszenciális rendjének felszálló ága véges (egy esszenciális okozat

gadhatónak Anzelm érvét: *Deus est quo cognito sine contradictione maius cogitari non potest sine contradictione* (*Ord.* I d. 2 p. 1 q. 1–2 n137, vö. uo. n11; n35–37; n138–139).

kiváltása az esszenciális okok *szimultán* fennállására utalt), míg az akcideniális rendben okok és okozatok végtelen *szukcessziója* valósul meg. Scotus tehát nem tagadja, hogy a kategoriáknak megfelelően rendezett világban adott a végtelen számú individuum lehetősége, ám nem fogja megengedni az „esszenciálisan rendezett okok” végtelen regresszusát. Az „esszenciálisan rendezett okok” sorozatának elkülönítése az „akcideniálisan rendezett okok” sorozatától akár arisztotelészi eredetű is lehetne,<sup>19</sup> Scotus azonban jelzi, hogy az általa bevezetett megkülönböztetés nem esik egybe Arisztotelész okfelfogásával. Valóban, beszélhetünk Arisztotelész szellemében arról, hogy vannak *per se* és akcideniális okok aszerint, hogy egy individuum lényegi vagy járulékos sajátossága szerint oka-e az okozatának.

n47 Más dolog *per se* és *per accidens* okról beszélni, és más dolog *per se* [azaz *essentialiter*] és *akcideniálisan rendezett* okokról beszélni. Az első esetben individuális dolgokról van szó, és arról, hogy egy dolog mint ok saját természetéből fakadóan ok, nem pedig valamely akcidense tekintetében. A második esetben két okot vetünk össze, amennyiben ugyanazon ok kiváltásában szerepet játszik mindkettő (Ed. Vat. II, 153–154. *Ord.* I d. 2 p. 1 q. 1–2 n47).

Az esszenciálisan rendezett ok és okozat n50-ben foglalt kategoriális különbségét többféleképpen értelmezhetjük. Egyfelől az egy nemen belüli kategoriális különbségekről lehet szó: az apa azon sajátossága, hogy ember, a nemzés viszonyában oka az ember sajátossággal rendelkező fiú létrejöttének – az apa–fiú viszonyban Scotus szerint kontingens módon rendezett ok és okozat lép fel, ám az apában fennálló *ember* sajátosság esszenciális oka a *fiúnak*. Scotus azonban nyitva hagy más, a tradicionális felfogástól sokkal nagyobb mértékben elütő lehetőségeket is. A szöveg alapján elgondolható ugyanis egy olyan kauzális viszony is, melyben az ok és az okozat radikálisan eltérő kategoriális berendezkedést mutatnak. Az egzisztenciálisan megvalósult dolgok körében lezajló kauzális viszony olyannyira önálló a dolgok kategoriális struktúrájával szemben, hogy ez utóbbi semmiféle elméleti kényszert nem ró az előbbire. Egy okozati hatásban természetesen Scotus szerint is formális sajátosságok (jellemzően fizikai minőségek) tranzíciója következik be, mely az ok és az okozat bizonyos kategoriális egyezését előfeltételezi, Scotus ezen elvében azonban benne rejlik annak lehetősége is, hogy bizonyos kvantitatív relációk megvalósulása kvalitatív változást idézzen elő. Annak érzékeltetésére, hogy Scotus milyen radikálisan tér el itt a létmeghatározottságok hierarchikus rendje és a kauzális viszonyok rendje közötti párhuzam koncepciójától, a *De primo principio* 2. fejezetének 13. és a 14. *conclusióját*

<sup>19</sup> *Fizika* 195a27–b3.

hozhatjuk fel. Ezekben a skót filozófus arról értekezik, hogy egy alacsonyabb rendű kategoriális berendezkedésű dolog okként léphet fel egy magasabb rendű kategoriális struktúrával bíró okozat szempontjából:

*Nem minden meghaladott (excessum) függ esszenciálisan a kiemelkedőtől (ab eminente). [...] A nemesebb faj kiemelkedő (est eminens) a kevésbé nemes faj szempontjából (mint például a nagyobb ellentét a kisebb szempontjából), és mégsem lép fel valamiféle okként arra nézvést. Ez induktív módon nyilvánvaló, amennyiben [a nemesebb] nem közelebbi okozata valamely közös oknak, mert a közös ok kauzalitása rájuk mint okozatokra nem az esszenciális rend szerint vonatkozik, mert ha ily módon vonatkozna rájuk, akkor csakis úgy volna képes valamely meghaladót okozni, hogy előbb a kiemelkedőt okozza, ami bármely okra nézve hamis. Ha ugyanis az alsóbb rendű ellentétet oly módon hozza létre ezen ok, hogy a felsőbbrendű ellentétet nem hozza létre semmiféle ok sem, akkor egyetlen ok szempontjából sem rendeződnek egymáshoz esszenciálisan. Végül, ha valamely kiemelkedő nem lép fel okként a meghaladott szempontjából, akkor nem is legközelebbi okozata kettejük [a kiemelkedő és a meghaladott] okának. Tehát a meghaladott nem függ attól esszenciálisan (Duns Scotus 1639. 219. concl. 13).*

A *De primo principio* terminológiájában a dependencia szigorúan kauzális függést jelent, míg a meghaladott (*excessum*) és a kiemelkedő (*eminens*) a kategoriális rend hierarchikus pozícióit jelentik. A tétel megengedi, hogy a magasabb rendű kategoriális meghatározottságú dolog okként lépjen fel egy vele univok, ám alacsonyabb rendű kategoriális szerkezettel bíró okozat szempontjából. A tézis negatív megfogalmazása azonban azt mondja ki, hogy a kauzális viszonyokkal szemben egyáltalán nem kell felállítanunk a kategoriális meghatározottságok párhuzamosságának e követelményét. Bármennyire is úgy tűnik, hogy a *Liber de causis* 2. tételéhez hasonló dolgot állít Scotus, a különbség mégis döntő jelentőségű. Egy magasabb kategoriális meghatározottsággal rendelkező dolog közvetítő kauzalitásának kiesése mellett nem azért állhat be mégis az okozat, mert egy annál még magasabb rendű primér ok kauzális hatása kiválthatja azt. Sokkal inkább arról van szó, hogy kategoriálisan egymáshoz képest teljesen rendezetlen okösszefüggések is vezethetnek ugyanazon okozat kiváltásához. Az azonos nembeli létmeghatározottságok közötti „fokozatkiesés” ténye kifejezetten amellet szól, hogy kategóriaugrás következhet be egy és ugyanazon okozat két különböző kauzális összefüggés következtében történő fellépésekor, még egyszer: *Ha ugyanis az alsóbb rendű ellentétet oly módon hozza létre ezen ok, hogy a felsőbbrendű ellentétet nem hozza létre semmiféle ok sem, akkor egyetlen ok szempontjából sem rendeződnek egymáshoz esszenciálisan.*

*Nem minden kauzálisan függő dolgot (dependens) halad meg (excessum) az, amitől függ (a quo dependet).* Nyilvánvaló, hogy az összetett szubsztancia kauzálisan függ az anyagtól, miközben jóval tökéletesebb annál. Hasonlóan: talán a forma is függ az anyagtól [...], azonban a forma tökéletesebb annál [...]. A rendezett változások esetében a keletkezés szerint későbbi függ az előbbitől, mert van egy mindkettőre nézve közös ok, és az utóbbi mégis tökéletesebb (Duns Scotus 1639. 219. concl. 14).

A 14-es konklúzió egyértelműsíti a 13-ban foglaltakat: az alacsonyabb kategoriális meghatározottság okként léphet fel a magasabb szempontjából az egzisztensen megvalósult dolgok szférájában. A rendezett változások arisztotelészi példájában nem annyira az antik filozófusnak a *dinamisz* fogalmában kifejtett természetfilozófiai princípiuma a hangsúlyos, mint inkább az a struktúra, hogy az ok és okozat azért pozícionált előbbi- és utóbbiként időben, mert van egy mindkettőt megelőző közös ok. E sorban a legutolsó lehet tökéletesebb, mint a második láncszem (lásd *Met.* 1050b2 skk.), azonban nem az arisztotelészi aktualitás eleve feltételezett magasabb metafizikai rangja miatt, hanem azért, mert az egzisztens kauzalitás kibújik a kategoriális possibilitás kényszerzubbonya alól.

Scotus a kauzalitás fenti törvényei alapján az esszenciálisan rendezett okok végtelen sorozata ellen érvel.

n52 Ezek alapján mutatjuk meg a feltevést, vagyis hogy

(A) az esszenciálisan rendezett hatóokok sorozatában lehetetlen a végtelen. [...]

(B) az akcideneciálisan rendezett hatóokok sorozatában csak akkor lehetséges a végtelen, ha az esszenciálisan rendezettek esetében nem mehetünk a végtelenbe. Tehát minden módon lehetetlen a végtelen regresszus az esszenciálisan rendezett okok rendjében.

(C) Sőt, még ha tagadnánk is azt, hogy létezne ilyen esszenciális rend, a végtelenség akkor is lehetetlen volna.

(Duns Scotus: Ed. Vat. II, 156–157. *Ord.* I d. 2 p. 1 q. 1–2, n52. A szakasz tagolása tőlem – S. J.)

(A) bizonyításához Scotus öt érvet hoz fel. Elsőként az esszenciálisan rendezett okozatok összességének (*universitas causatorum essentialiter ordinatorum*) eszméjére utal. Az esszenciális okozatok összességének oka nem lehet okozat, hiszen akkor ezen ok is része volna az esszenciális okozatok összességének, vagyis ezen összesség oka önmagán belül volna fellelhető. Az okozatok összessége nem lehet *causa sui*. Az ötödik érvet idézzük hosszabban.



n53 Ötödször: a hatóokság szükségszerűen semmiféle tökéletlenséget nem foglal magába, így képes benne lenni valamiben tökéletlenség nélkül.<sup>20</sup> Ha azonban semmiféle olyan ok nem létezne, amely ne függne valami öt megelőzőtől, akkor semmiben sem volna meg tökéletlenség nélkül. Tehát a független hatóokság (*effectibilitas independens*) képes benne lakozni valamely természetben, és ez a *simpliciter* első: tehát lehetséges az első hatóokság (*effectibilitas*). Ez kielégítő, mert majd ez alapján következtetünk,<sup>21</sup> mivel ha egy ilyen első hatóok (*efficiens*) lehetséges, akkor valóságosan is létezik (uo. 158–159).

Scotus ötödik érve azt mondja ki, hogy az esszenciális kauzalitás minden okozat viszonyban tökéletességet mutat akkor is, ha az ok maga is okozat az esszenciális rendben. E tökéletesség az n50-ben rögzített kategoriális különbségből adódik. Másfelől az ok tökéletessége a kauzális viszonyban az okozat felé nyilvánul meg, ő maga mint valamely megelőző ok esszenciális okozata, bizonyos tökéletlenséggel bír. Scotus arra mutat rá, hogy lehetséges egy olyan ok, mely híján van minden tökéletlenségnek.

Ez az egyik olyan érv, amely jól érzékelteti Scotus azon erőfeszítését, hogy a formális sajátosságok kategoriális rendjét – melyet aprólékos műgonddal tisztított meg a kauzalitás befolyásától – mégiscsak beleágyazza a teremtett világ ok–okozati összefüggésrendszerébe. Erről tanúskodik óvatossága is: a hatóoki viszonylatok evilági diagnózisától nem ugorhatunk mindjárt egy első hatóok tételezéséig, ahogy azt Aquinói Tamás teszi. Scotust sokkal jobban érdekli az az általa közbevetett lépcső, amit az első hatóok intelligibilitásából fakadó lehetőség jelent. A kauzalitás tökéletességének intelligibilis lehetősége a kulcsmomentum az első hatóok egzisztens létezéséhez. A kauzalitás intelligibilitásának kérdése nem más, mint a kategoriális rend – az egyáltalában vett intelligibilitás – peremvidéke elgondolhatóságának kérdése, vagyis: hogyan lehetséges az egzisztens létezés szférája, azaz hogyan lehetséges a nem-pusztán-lehetséges (hanem valóságos) területe? Nem annyira Scotus válasza érdekes, mint inkább a megközelítésmódja: az egzisztens valóság legfőbb jellegzetességét, azaz a kauzalitást is annak lehetősége, intelligibilitása felől közelíti meg. Scotus szerint lehetséges a kauzalitás tökéletessége. *Sed hoc sufficit*: ez már elégséges alapot fog

<sup>20</sup> *Potest esse in aliquo sine imperfectione*. Filozófiatörténeti érdekességként megjegyezhetjük, hogy a 16. századi kései skolasztika idején az *omnis perfectio sine aliqua imperfectione* scotusi gondolata kifejezetten ellentmondásba kerül a *primum movens immobile* gondolatával. Egy Christian Francken nevű német gondolkodó 1590 körül Kolozsvárott papírra vetett művében úgy vélte, hogy ha a scotusi legfőbb tökéletességű első hatóokot nem lehet első mozgatatlan mozgatóként tételezni, akkor a filozófiai teológia egyetlen konzekvens álláspontja az ateizmus.

<sup>21</sup> Duns Scotus, *Ed. Vat.* II, 164–165 (*Ord.* I d. 2 p. 1 q. 1–2, n58, lásd lent).

nyújtani ahhoz is, hogy e lehetőségből később az első hatóok egzisztens létezésére következtesen.

A kauzalitás tehát, mely legsajátabb természeténél fogva a kategoriális rendentől – intelligibilis tartalmát tekintve mintegy a kategoriális rendtől függetlenül – jellemzi a világ dolgait, ugyanúgy bizonyos intelligibilis tartalommal rendelkezik, mint a világ egzisztens dolgainak más kategoriális sajátosságai. E sajátosságok esetében elkülöníthetjük ezek formálonológiai és egzisztens-kauzális szempontú megközelítéseit. Formálonológiai szempontból az egzisztensen inhereáló sajátosság fennállása önnön *formális lehetőségéből* következik: az aktuálisan létező dolog sajátosságalmazásával nem áll ellentmondásban (Scotus pozitív terminológiájával: *compossibilis*), hogy a szóban forgó sajátosság megjelenjen a dolgon. E fennállás – függetlenül attól, hogy milyen kauzális relációban valósul meg az adott dolog – a sajátosság possibilitása, az ellentmondásmentesség elve felől artikulálódik, így formálisan *a se* fennállás.<sup>22</sup> Egzisztens-kauzális szempontból viszont a szóban forgó sajátosság intelligibilis tartalma önmagában nem elég annak megvalósulásához, ezért e sajátosság megvalósulása nem *a se* létezés. Mit mondhatunk azonban a kauzalitásról akkor, ha elfogadjuk a Scotus által bejelentett igényt arra vonatkozóan, hogy a kauzalitás intelligibilis tartalomra tehető?

n58 Az, aminek elvével ellentmondásban áll az, hogy más által legyen, ha képes létezni, akkor képes önmaga által létezni (Duns Scotus: Ed. Vat. II. 164. *Ord.* I d. 2 p. 1 q. 1–2 n58).

A kauzalitás intelligibilis tartalma abban fejeződik ki, hogy általa egy olyan mozzanatot vezetünk be az intelligibilitás területére, ami épp az abból való „kilépést” – vagyis a puszta lehetőség szférájának elhagyását – hivatott biztosítani. A kauzalitásnak mint *elgondolt* sajátosságnak így pontosan a mástól való tényleges dependencia jellegzetességét kell maga mögött hagynia. A tökéletes kauzalitás intelligibilis tartalma, a „mástól való függőség” lehetősége (az *ab alio* possibilitása) a tényleges fennállástól elválasztva is megőrzi a kauzalitás formális szerkezetét: ez lesz az *aseitas* (az önmagától eredő jelleg) a kauzalitás formális maximumában. Ám nemcsak arról van szó, hogy a kauzalitás fogalmi tartalma formálisan, az egzisztens létezés tekintetében indifferens módon, önmagában megalapozható (formálonológiai *aseitas*) – ez Scotus szerint minden általános sajátosságra jellemző. A kauzalitás abban tér el más sajátosságoktól, hogy elgondolhatóságából következő tartalma maga az *aseitas*. Míg a „fehér” sajátosság esetében ez a tartalom pontosan a *fehér* (minőség, szín stb.) kategoriális jellemzőiben merül ki, addig az okság esetében egy olyan ismerettartalomról van szó, mely a kategoriális

<sup>22</sup> A dolog sajátosságának formális lehetősége Scotus szerint képes kiváltani a rá vonatkozó kognitív aktusunkat: így ismerjük meg a *reálisan* önzonos dolgokon a gyenge, azaz *formális* identitású intelligibilis sajátosságokat.

rend bármely más tagjától eltérően nem szorul rá az egzisztens létszféra kauzális összefüggésrendjére ahhoz, hogy megvalósuljon.

Scotus e gondolatmeneteinek célja az első hatóok egzisztens létezésének bizonyítása. E bizonyítás a kauzalitás mint magyarázó elv formális természetének vizsgálatából indul ki, és az így kapott reduktív kauzalitásfogalom puszta intelligibilitása alapján bizonyítja isten létezését. Az istenérv e formájának talán az a legfontosabb tanulsága, hogy a teremtés tézise, isten világra vonatkozó kauzalitásának koncepciója fenntartható akkor is, ha a teremtett dolgok sajátosságainak lehetősége függetlenedik e kauzális viszonyról. A *fehér* sajátosság intelligibilis tartalma akkor is ugyanaz volna, ha isten nem teremtette volna meg a világot. Ennek ellenére Scotus istent a teremtés kauzális viszonyában kontingens okként állította be, aminek azonban ára volt: az, hogy a világ lehetne másképp is, a possibiliis, a létező mint létező fogalma alatt ellenmondásmentesen elgondolható világok gondolatát feltételezte; tehát pontosan a kauzalitás mint magyarázó elv hatókörének háttérbe szorulását. Hogy ezen ellentmondásmentességnek az isteni mindenhatóságra vonatkozó kényszerével Scotus hogyan számol el, az természetesen egy másik kérdés.

## IRODALOM

- Avicenna 1992. *Avicenna Latinus. Liber primus naturalium de causis et principiis naturalium*. Szerk. Simone Van Riet – Gérard Verbeke. Louvain-La-Neuve – Leiden, Peeters Publishers.
- Borbély Gábor 2008. *Cirkulóó angyalok. Bevezetés a középkori filozófiába*. Budapest, Akadémiai Kiadó.
- Duns Scotus, Ioannes 1639. *De primo principio*. In Luke Wadding (szerk.) *Ioannis Duns Scoti Opera Omnia*. Lyon, Laurentius Durand. Vol. 3.
- Duns Scotus, Johannes 1891–1895. *Ioannis Duns Scoti Opera omnia*. Paris, Louis Vivès (= Ed. Viv.).
- Duns Scotus, Johannes 1950. *Ioannis Duns Scoti Opera omnia*. Studio et cura Commissionis Scotisticae ad fidem codicum ed. praes. Carolo Balić. Civitas Vaticana, Typis Polyglottis Vaticanis (= Ed. Vat.).
- Duns Scotus, John 1966. *A Treatise on God as First Principle*. Ford. Allan B. Wolter. Chicago, Franciscan Herald Press.
- Duns Scotus, Johannes 1974. *Abhandlung über das erste Prinzip – Tractatus de primo principio*. Szerk. Wolfgang Kluxen. Darmstadt, Wissenschaftliche Buchgesellschaft.
- Duns Scotus, John 1994. *Contingency and Freedom. Lectura I d. 39*. Ford. és szerk. Vos Antonie Jaczn et alii. Dordrecht–Boston–London, Kluwer.
- Duns Scotus, Johannes 1997. *Ioannis Duns Scoti Opera Philosophica*. Szerk. Girard J. Etzkorn. St. Bonaventure, New York, Franciscan Institute (= Op. Phil.).
- Honnefelder, Ludger 1990. *Scientia transcendens – Die formale Bestimmung der Seiendheit und Realität in der Metaphysik des Mittelalters und der Neuzeit (Duns Scotus – Suárez – Wolff – Kant – Pierre)*. Hamburg, Felix Meiner Verlag.
- Huoranszki Ferenc 2001. *Modern metafizika*. Budapest, Osiris.
- Knuuttila, Simo 1993. *Modalities in Medieval Philosophy*. London, Routledge.
- Liber de causis* 1966. Szerk. A. Pattin. In *Tijdschrift voor Filosofie* 28. 90–203.

- Perler, Dominik 2002. *Theorien der Intentionalität im Mittelalter*. Frankfurt am Main, Klostermann.
- Perler, Dominik – Ulrich Rudolph 2000. *Occasionalismus. Theorien der Kausalität im arabisch-islamischen und im europäischen Denken*. Göttingen, Vandenhoeck & Ruprecht.
- The Book of Causes [Liber de Causis]* 1984. Ford. és szerk. Denis J. Brand. Milwaukee, Marquette University Press.
- Thomas de Aquino 1965. *Sancti Thomae de Aquino Expositio super librum Boëthii De Trinitate*. Szerk. Bruno Decker. Leiden, Brill.
- Vos Jaczn, Antonie 2006. *The Philosophy of Duns Scotus*, Edinburgh, Edinburgh University Press.
- Zimmermann, Albert 1965. *Ontologie oder Metaphysik, die Diskussion über den Gegenstand der Metaphysik im 13. u. 14. Jahrhundert*. Köln–Leiden, Brill.

## Arisztotelészi-skolasztikus elemek a descartes-i szenvedélytanban

### I. BEVEZETÉS

A test–lélek kapcsolat descartes-i felfogását különféle irányokból közelíthetjük meg. A potenciálisan tárgyalható témák egy része a pszichofizikai okság *metafizikai feltételeire* vonatkozik. Feltehetünk például olyan kérdéseket, hogy annyira különböző természetű szubsztanciák, mint a test és a lélek, előidézhetnek-e hatást egymásban, vagy hogy a fizikai világ mint zárt rendszer szenvedhet-e egyáltalán külső ráhatást. Ezeket a – gyakran vitatott – problémákat ehelyütt félre tesszük, hogy az okság metafizikai feltételei helyett az oksági folyamatok *tényleges lezajlásának* lényegesen kevésbé feltárt részleteivel foglalkozzunk.

Az utóbbi téma viszonylagos népszerűtlensége részben a tobozmirigy-hipotézis általánosan elismert abszurdításával hozható összefüggésbe. Szintén hozzájárulhatott ehhez az a tény, hogy a szorosan vett pszichofizikai eseményekre vonatkozó descartes-i leírások szűkszavúak és igen homályosak. Sok szó esik arról – például a *Traité de l'homme* és a *Les passions de l'âme* lapjain –, hogy mi történik a testben addig a bizonyos pontig, ahol a két szubsztancia egymásra hat. Magát a pszichofizikai átmenetet ezzel szemben Descartes, mint látni fogjuk, egy-egy nehezen értelmezhető kifejezéssel intézi el, amelyek ráadásul egymással sem tűnnek összeegyeztethetőnek.

A továbbiakban értelmezni próbálom a külső és belső érzékelésben, valamint az érzelmekben megjelenő pszichofizikai eseményeket. Arra a megállapításra fogok jutni, hogy a test–lélek interakció descartes-i leírása csak az érzékelésre vonatkozó arisztotelészi-skolasztikus felfogás(ok) ismeretében érthető meg. Ennek nem csupán az az oka, hogy Descartes a hagyományos érzékelésmélelet kifejezett kritikával illeti, hanem véleményem szerint maga a descartes-i elgondolás is – s nem is csupán a szerző korai műveiben – a bírált nézetek elemeire épül.

## II. DESCARTES ÉRZÉKELÉSELMÉLETE: HOGYAN MEGY VÉGBE A PSZICHOFIZIKAI KÖLCSÖNHATÁS?

A karteziánus érzékelésmélet tárgyalását a *Regulae ad directionem ingenii* című, az 1620-as években keletkezett műben található leírással kezdem. A tizenkettedik szabály szerint:

Először is, amennyiben a külső érzékszervek a test részei, az érzékelés szigorúan véve szenvedés, még akkor is, ha az érzékszervek odafordítása a dolgokhoz cselekvés, azaz mozgás; ugyanolyan módon zajlik, ahogyan a viasz felveszi a pecsétnyomó alakját. S nem szabad azt gondolnunk, hogy ez csupán analógia: az érzékelő test külső alakja (*figuram externam corporis sentientis*) valóságos változást szenved a tárgytól, pontosan olyan módon, ahogy a pecsétnyomó változást idéz elő a viasz felszínén (AT X. 412).<sup>1</sup>

Mindez, állítja Descartes, a többi érzékre is igaz, amelyek mind befogadnak valamely formát vagy alakot (*recipere figuram impressam*). A pecsétnyomó-analógia Arisztotelésztől származik, akinek *De anima* című munkájában ezt olvassuk: „az érzék az, ami az érzékelhető formákat anyaguk nélkül befogadni képes, ahogy a viasz a pecsétgyűrű jelét a vas vagy arany nélkül fogadja magába [...] Hasonlóképpen minden egyes érzékelhető minőség esetében érzékelésünk attól szenved el hatást, aminek színe, íze, zaja van” (II 12, 424a17; Arisztotelész 2006. 63). Ami a hivatkozott *figurát* (formát vagy alakot) illeti, Descartes megtartja az arisztotelészi szóhasználatot, amennyiben ezek is „tisztán”, „anyag nélkül” (AT X. 414) érkezik az érzékszervek felől. Az érzékelés Descartes számára is *változás*, valamilyenné válás, melynek során az elme egy új módosulásra tesz szert. Descartes azt az arisztotelészi gondolatot is megőrzi, hogy az érzékelésben foglalt ráhatás és az elszenvedés (*actio* és *passio*) egy és ugyanazon aktus, két különböző nézőpontból szemlélve (vö. AT XI 328). S amint a pecsétnyomó-analógiából már világossá vált, a – valamilyen értelemben vett – forma passzív átvétele is arisztotelészi elem, csakúgy, mint a közös érzék (*sensus communis*) szerepeltetése.

Amiben az érzékelés descartes-i leírása leginkább különbözik attól az elmélettől, amelyet Descartes az arisztotelészi-skolasztikus hagyománynak tulajdonít, az az, hogy szerinte az érzékelés folyamata során nem kerül át semmiféle entitás a lélekbe. Descartes érzékelésméleteének későbbi részletei leginkább e gondolat köré összpontosulnak. Descartes tagadja a skolasztikus *speciesek* létjogosultságát, s – tágabban – azt, hogy az érzékelés a hasonlóság elvén alapulna; az utóbbi helyébe egyfajta izomorfizmus lép az érzéki minőségek és az azokat az

<sup>1</sup> A rövidítés feloldását lásd a tanulmány végén az irodalomjegyzékben. Ahol másként nem jelzem, ott a fordítások tőlem származnak.

elme számára közvetítő testi mozgások között. A folyamat keretelképzelése és a pszichofizikai interakció felfogása azonban – s ez dolgozatom fő állítása – nem sokat változik.

Térjünk rá tehát magára a pszichofizikai mozzanatra. A *Regulae* koncepciója szerint az érzékelés ugyanúgy történik, ahogyan a viasz felveszi a pecsétnyomó alakját, tehát az elme szempontjából alapvetően passzív folyamat, annak ellenére, hogy – ahogyan az a fenti idézetben szerepel – az érzékeket *odafordítjuk* (*applicemus*) a tárgyakhoz, ami cselekvést jelent. Ez az odafordulás nem teszi magának az érzékelés folyamatának a során aktívvá az elmét, hiszen annak csupán mintegy a kezdeményezése. Van azonban egy másik „odafordulás”, tudniillik az, amely magát az érzékelést („látást, tapintást stb.”) *jelenti*, amellyel felfogóképességünk (*vis cognoscens*) „a képzelőerővel együtt” a közös érzéket, illetve az általa tárolt információt illeti (AT X 415). A felfogóképesség a különféle mentális funkciók gyakorlása során mindig valamely *figura* felé fordul (lásd fent), vagy azt szemléli (AT X 415–416). Descartes nem hagy kétséget afelől, hogy ez az odafordulás a passzív folyamat része, melynek során a lélek hatást szenved el, s a viasz, nem pedig a pecsétnyomó módján működik. Mivel „a testi dolgokban nem találunk ilyesfajta képességet”, a felfogóképesség aktusa és a pecsétnyomás között csupán *párhuzam* áll fenn (vö. AT X 415).

A *Traité de l'homme*-ban részletesebb kísérlettel találkozunk a látás magyarázatára. A külső tárgyak felületéről érkező fénysugarak, írja itt Descartes, nyomást gyakorolnak a szemfenékre, ahonnan idegek nyúlnak az agy belsejéig (vö. AT XI 175). Amikor a szemfenék valamely pontját nyomás éri, a hatás az idegroston keresztül továbbítódik, s az agyban kitágul egy vezeték, melyen keresztül életszellemekek indulnak el nagyobb számban a tobozmirigy felől. Így ugyanaz a forma rajzolódik ki a szemfenéken, az agy belső felületén, majd a tobozmirigyen. Ez a folyamat anélkül zajlik, hogy bármiféle dolog ténylegesen átkerülne a tárgyról az érzékszervekhez, illetve az agyhoz. Ami számunkra leginkább lényeges, az az, hogy e leírás szerint a lélek, amikor elképzeli vagy érzékeli valamit, „közvetlenül szemléli” a tobozmirigy felületén megjelenő alakokat, formákat vagy képeket. Nem kapunk rá magyarázatot, hogy az utolsó fázis ezen eseménye, melynek során a lélek mintha aktív szerepet játszana, hogyan zajlik, és az érzékelés hogyan lehet ettől még a lélek szempontjából passzív.

A *Traité de l'homme* gondolatmenetében természetesen ismét szerepet kapnak a formák vagy alakok, melyekkel a *Regulae*-ban találkoztunk.

És figyeljenek fel arra, hogy ezen alakok alatt nem csupán azokat a dolgokat értem, amelyek valamilyen módon reprezentálják a tárgyak vonalainak és felületeinek elhelyezkedését, hanem mindazokat, amelyek a fentebb mondottaknak megfelelően képesek alkalmat adni a léleknek arra, hogy érezze a mozgást, a nagyságot, a távolságot, a színeket, a hangokat, a szagokat és más hasonló minőségeket; de ugyanígy azokat is, amelyek képesek a

csiklandozás, a fájdalom, a szomjúság, az öröm, a szomorúság és más hasonló szenvedélyek érzésére bírni a lelket. [...] Ezen alakok közül nem azokat tekintjük ideáknak, amelyek a külső érzékek szerveibe, de nem is azokat, amelyek az agy belső felületére nyomódnak. Ideáknak, tehát olyan formáknak vagy képeknek, amelyeket az értelmes lélek közvetlenül szemlél, ha, egyesített lévén e géppel, elképzeli vagy érzékel valamit, csakis azokat az alakokat nevezzük, amelyeknek nyomait az életszellemek *a képzelet és a közös érzék székhelyéül* szolgáló [toboz]mirigy felületén alakítják ki.

Figyeljenek fel arra, hogy amikor azt állítom: a gép „elképzeli” vagy „érezkelet”, azért teszek így, mert az idea neve alá általában be akarom foglalni mindazokat a benyomásokat, amelyeket a [toboz]mirigyből kilépő szellemek képesek magukba fogadni; e benyomások közül pedig azok mindegyike a közös érzéknek tulajdonítható, amelyek a külső tárgyak jelenlététől függenek; azonban számos más okból is előállhatnak, amint ezt az alábbiakban kifejtem, és ekkor a képzeletnek kell tulajdonítanunk őket (AT 177, magyarul: 39).<sup>2</sup>

A metafizikai-episztemológiai orientációjú Descartes-kutatás számára az itt tárgyalt gondolatmenet leglényegesebb eleme az volt, hogy Descartes a tobozmirigyen megjelenő képeket *ideáknak* nevezi. Az ideák, mint tudjuk, később tisztán mentális jellegűekké válnak Descartes-nál. Ez a kettőség számunkra azért figyelemre méltó, mert jelzi Descartes küzdelmét azzal az örökölt problémával, hogy az érzékelésből származó tartalom hogyan válhat a „racionális lélek” által szabadon manipulálhatóvá és olyan esetekben is hozzáférhetővé, amikor külső tárgy nincsen jelen, oly módon, hogy közben megőrzi azonosságát. A Descartes által elvetett hagyomány megoldása (többek között) erre a problémára az érzéki speciesek intelligibilis speciesekké (majd fogalmakká) történő konverziója. Descartes-nál a mentális reprezentációk az egységes elme bármely művelete számára közvetlenül hozzáférhetőek, az érzéki tartalmak a „közvetlen szemlélésben” válnak ilyenekké. Bár a megnevezésben, mint látjuk, Descartes kissé bizonytalanul látszik, az arisztoteliánus-skolasztikus szerzőkhöz hasonlóan arra törekedett, hogy magyarázatot adjon egyrészt arra, hogy az érzéki információ fiziológiai változásokon keresztül miként adódik át a megismerő elmének, másrészt arra, ahogyan az utóbbi saját jellegének megfelelően kezeli ezt az információt.

Aquinói Tamás egyaránt beszél az érzéki speciesek megragadásáról és befogadásáról (vö. Stump 2003, 249–250). Az előbbi megfogalmazás azt az értelmezést sugallhatja, hogy maga a species az, amit Tamás szerint érzékelünk. Tamás viszont azt állítja, hogy az érzéki species nem az, amit érzékelünk, hanem az, *aminek a segítségével* érzékelünk (uo. 250). A forma, melyet az érzéki species

<sup>2</sup> A *Traité de l'homme* szövegét Gulyás Péter megjelenés előtt álló – helyenként módosított – fordításában idézem a kézirat oldalszámára utalva.



továbbít, az alak, a térbeli elhelyezkedés és az oksági viszonyok egyfajta konfigurációjának tekinthető (uo. 36), s ez nem áll távol attól a descartes-i „alaktól”, amely a tobozmirigyen megjelenik. Maga az „alak” Descartes számára sem közvetlenül hozzáférhető az érzékelésben és a gondolkodásban, hanem *érezkellet* valamit az alannal. (Az érzékelés *valamilyen dologként* való érzékelés lesz, amihez szükség van az elme további munkájára – lásd alább, az érzékelés három lépcsőfokáról.) Az érzéki tartalmak közvetítői tehát mint ilyenek maguk nem tárgyai az érzékelésnek, de a tényleges érzéklet ezekre épül.

A speciesek átadásának fázisai már egyes skolasztikus szerzőknél átalakultak önállóan értelmezhető fizikai mozgások sorozatává. Suareznál például, aki megtartja a species-terminológiát, a species *in medio* (vagyis a közegben zajló) fizikai változások egymásutánját jelenti, s Suarez ezzel összefüggésben (sok más szerzővel együtt) a specieseknek az érzéki tárgyakhoz hasonló képek módjára történő felfogását is elveti (vö. Aho 2007. 188–189). Innen már csak egy lépés, a mechanisztikus szemlélet hozzáadása szükséges ahhoz, hogy az érzékelést a tárgyak olyan hatásának tekintsük az érzékelőre, amelyet a testben levő részecskék mozgása közvetít izomorf „alakok” létrehozásával.

De térjünk vissza az egyes Descartes-művek érzékeléstanára, illetve Descartes ellenvetéseire az általa ismert magyarázatokkal szemben. Az 1630-as évek elején keletkezett *Dioptrique* tömör – és némiképp igazságtalan – kritikáját adja a skolasztikus érzékeléstanak:

Óvakodnunk kell attól, hogy feltételezzük – amint filozófusaink általában teszik –: ahhoz, hogy a lélek érzékelni tudjon, képeket kell szemlélnie, amelyek a tárgyakról átadódnak az agynak; vagy legalábbis ezen képek természetét egészen másképpen kell felfognunk, mint a filozófusok. Mivel ugyanis az ő elképzelésük ezekről a képekről arra a követelményre korlátozódik, hogy azoknak hasonlítaniuk kell a tárgyakra, melyeket megjelenítenek, nem képesek megmutatni, hogy a tárgyak hogyan hozzák létre a képeket, hogyan fogadják be azokat a külső érzékszervek, és hogyan kerülnek át az idegeken keresztül az agyhoz (AT VI. 112).

Descartes tehát egy olyan nézetet állít itt pellengérré, amelyet korántsem mindenki fogadott el. Ami saját nézeteit illeti, a *Dioptrique*-ban úgy vázolja a pszichofizikai interakciót, hogy az agyban zajló mozgásokból álló minta „közvetlenül hat a lélekre”, és a természet által elrendelt módon (AT VI. 130) kiváltja benne az érzést. A folyamat vége ehelyütt inkább passzívként jelenik meg: „[F]el kell tételeznünk, hogy lelkünk olyan természetű, hogy ami kialakítja vele a fény érzését, az *azoknak a mozgásoknak az ereje (la force des mouvements)*, melyek az agy azon részein zajlanak, ahol az optikai idegrostok erednek” (uo.).

Burman a Descartes-tal folytatott beszélgetés során kifejezetten rákérdez a pszichofizikai interakcióban foglalt mentális aktus mibenlétére:

Mit jelent, hogy „szemléli” [tudniillik a testi állapotot]? Ugyanazt jelenti, mint a „megértés”? Ha igen, miért használ rá más kifejezést? Ha nem, az elme nem csupán megértő és gondolkodó dolog lenne, és még mielőtt teste lenne, már képes szemlélni azt. Vagy az elme ezen képessége a testtel való egyesítettség hatása? (AT V. 162.)

Descartes válasza nem túlságosan informatív. A szemlélés, melyről szó van, „a gondolkodás speciális modifikációja” (s így a testtel való egyesítettségből ered). Mi tehát ez a mentális aktus vagy állapot? Aligha vélhetjük, hogy külön intencionális aktusról lenne szó, amely az agy valamely állapotára irányul: Descartes szerint elménk minden állapotával tisztában vagyunk, és nincsen tudomásunk ilyen aktusokról. Az érzékelés folyamatának utolsó eleme továbbá – a folyamat egészéhez hasonlóan – alapvetően passzív. Ezt a vonását kell összeegyeztetni a ténnyel, hogy valamiféle „odafordulásról” és „szemlélő” tevékenységről van szó. A *figyelem* valamilyen tárgyra irányítása – amennyiben erről van szó – Descartes-nál az elme aktív működése (vö. AT XI. 361). Igaz, hogy az intellektus minden tevékenysége *perceptió*, s ennyiben passzív (vö. AT VIII-1. 17). Mivel azonban Descartes a tiszta értelem működését is ide sorolja, úgy tűnik, tág tere van az elme – legalábbis köznapi értelemben vett – *tevékenységének* egy definíció szerint passzív állapotban is. Érdeemes tehát felderíteni, milyen tevékenységről lehet szó a szenvedélyek esetében.

### III. A DEKÓDOLÓ ÉS A FELISMERŐ/KORRIGÁLÓ ELME

Descartes-nál a külső és belső érzékelés, illetve az érzelmek esetében olyan változások következnek be a tobozmirigyben, melyek révén az elme különféle érzetekre tesz szert. Az elme a descartes-i felfogás szerint ilyenkor passzív, s a test részéről hatást szenved el. Hogyan gyakorol egy test hatást valamely más dologra?

Amikor egy mozgó test egy másik testnek ütközik, mozgása a *Principia* harmadik természettörvénye szerint a másik test „erősebb” vagy „gyengébb” mivoltának függvényében változik.<sup>3</sup> (Legalább) két eseménnyel, illetve folyamattal van tehát dolgunk: a két test először összeütközik, majd megváltozott módon folytatja mozgását. Ha a pszichofizikai folyamatok ezzel analóg módon zajlanának le, ennek látszólag az lenne a módja, hogy a tobozmirigyben bekövetkező pszichofizikai esemény után, azzal koordinálva vagy annak hatására új érzékelésemény lép fel az elmében. Bizonyos értelmezők egyenesen az okság hume-i, két esemény állandó kapcsolatát tételező felfogásának megelőlegezőjét látják Descartes-ban. Louis Loeb szerint a descartes-i pszichofizikai okság abból

<sup>3</sup> Az első két törvény nem változásokat ír le, hanem azt, hogy mi történik azok hiányában.

állna, hogy bizonyos típusú lelki események bizonyos típusú fiziológiai eseményekre következnek, az akaratlagos mozgás esetében pedig fordítva (Loeb 1981. 137). Descartes – vélt vagy valós – okkazonalizmusának tárgyalását is gyakorta vezérli hasonló elgondolás: az lenne tehát a kérdés, hogy „testi események valóságos hatóokai-e lelki eseményeknek, illetve lelki események [...] valóságos hatóokai-e testi mozgásoknak” (Nadler 1997. 161).

A hume-i okságfelfogással való párhuzam ellen szól, hogy Descartes nem eseményeket szokott okoknak tekinteni. Az említett mozgástörvények esetében is Istent, illetve magukat a törvényeket nevezi meg mint okokat (s természetesen személyek, dolgok is lehetnek ágensek). Hogy nem *két* eseményről van szó, a *Passions* azon, már említett megfogalmazása is sugallja, mely szerint a ráhatás és az elszenvedés (*actio* és *passio*) egy és ugyanazon aktus, két különböző nézőpontból szemlélve. Hogyan „fordul” tehát a lélek bizonyos minőségek felé, hogyan „szemléli” az agy belsejében megjelenő mintázatokat úgy, hogy *ez egyszersmind* egy a testből származó hatás elszenvedése is?

Tovább bonyolítja a kérdést, hogy a testi és a lelki esemény, állapot vagy minőség közötti viszony nem mindig látszik *hatóoksági* viszonyának Descartes-nál. Descartes analógiát állít fel *jelölő* és *jelölt* viszonyával (sőt talán túl is megy az analógián):

[H]a a szavak, melyek csupán emberi konvenció révén jelentenek valamit, olyan dolgokat tudnak gondoltatni a lélekkel, melyekre nem hasonlítanak, a természet miért ne alakíthatott volna ki valamely jelet, amely a fény érzését keltené az elmében, akkor is, ha semmi sincs ebben a jelben, ami hasonlítana a fényre? (AT XI. 4.)

Yolton felhívja a figyelmet arra, hogy ha jelölési viszonyról van szó az érzékelésben, sokkal inkább azt várnánk, hogy az idea vagy érzet jelölje a külső dolgot. Ezzel szemben Descartes-nál, úgy tűnik, az agyon kirajzolódó minta jelöli az érzékletet, melynek kialakítására módot ad (Yolton 1996. 196). (A *reprezentáció* – nem pedig a *szignifikáció* – viszonya ezzel szemben az agyban bekövetkezett változások és a külső tárgy között állna fenn, lásd AT XI. 369.) Erre a Yolton által „megfordított jelölési viszonyt” nevezett struktúrára más szerzők is hivatkoznak, és a pszichofizikai esemény mentális elemét mint *dekódolást* jellemzik (lásd pl. Wilson 1993). Az ideák tehát nem dolgokat jelölnek, hanem fizikai változásokat értelmeznek, s az elmének e felfogásban az értelmező szerepe jut.

Az értelmezés, illetve a jelölés kontextusában mozog-e tehát a pszichofizikai esemény helyes interpretációja? Több érv szól az ellen, hogy egyszerű információátadásról vagy jelentésértésről van szó, mint azt elme „szemantikai” felfogás sugallná. Ha Descartes komolyan úgy gondolta volna, hogy a jelölési viszony számot ad az érzékelés folyamatáról, miért nem említette a különböző természetű szubsztanciák közötti *hatóokság* lehetőségéről folyó vitákban, hogy valójában

egészen másról, a jelölési viszonyban foglalt *formai* okságról van szó? Az általános benyomás, melyet Descartes-nak az érzékeléssel kapcsolatos kijelentéseiből nyerünk, nagyon is az, hogy a hatóokság lehetőségét óhajtotta bizonyítani, vagy legalábbis ezt tételezte fel még akkor is, ha ehhez annak fogalmát tágabban kellett értelmeznie, mint például a fizikai testek kölcsönhatása esetében.

Másrészt, ahogyan azt a *Hatodik elmélkedés* is megvilágítja, az elme nem csupán *megér*ti, hogy a test valamely ártalmat szenvedett el, mint a kormányos a hajójával kapcsolatban, hanem *érzi* azt. Annak, ahogyan az elme elszenvedti a ráhatást és befogad egy formát, sajátossága az a „színes-szagos” mód, ahogyan a világot érzékeljük. A dekódolás-értelmezés nem magyarázza az érzetek közvetlen voltát és az érzéki modalitások sokféleségét.

Harmadrészt, amint maga Yolton is elismeri, „helytelen azt mondani, hogy Descartes kvázi-személyt vagy második érzékelőt csinált volna az elméből” (Yolton 1996. 209). Descartes kifejezetten állítja, hogy az interakció mentális eleme nem az agyban kirajzolódó forma „leolvasásából” áll: „Mint már megmutattam ugyanis, nem szabad azt gondolnunk, hogy a kép e hasonlóság okán váltja ki ezen tárgyak érzékelését – mintha az agyban további szemek lennének, amelyekkel felfognánk azt” (AT VI. 130).

Tagadhatatlan, hogy a szemantikai viszony – inkább analógiaként, mintsem megoldási javaslatként – jelen van Descartes érzékeléselméletében. Célja azonban elsősorban az, hogy az érzékelésben foglalt hasonlóság tézise ellen szolgáljon érvként. A jelölés elmélete logikai-nyelvfilozófiai elmélet, mely közvetlenül aligha használható az érzékelés mechanikájának magyarázatára.

Egy másik nézet szerint az elme hozzájárulása az érzékeléshez az olyan aktusokban áll, mint például amikor figyelmünket a kezünkötől kifelé irányítjuk tárgyak helyzetének meghatározásánál, vagy amikor a perspektivikus torzításokat kiigazító ítéletet hozunk. A színlátás példájával élve: „A színeket térben kiterjedteknek érzékeljük, s így a színérzékelés összekapcsolódik a méret és az alak érzékelésével, amelyek viszont – Descartes nézetének értelmében – a helyzetével és a távolságával függnek össze. S mivel Descartes szerint a mozgások mintázata a tobozmirigynél szerkezetileg izomorf a retinán megjelenő képpel (amely kétdimenziós, fordított, méret és alak tekintetében perspektivikusan torzult), meg kell magyaráznia, hogyan érzékelhetjük helyesen a látás segítségével a tárgyak helyzetét, távolságát, méretét és alakját” (Wolf-Devine 2000. 510–511).

Wolf-Devine is felveti az elme passzivitásának problémáját az érzékelésben, amely látszólag nehezen egyeztethető össze azzal, hogy ítéleteket hoz alak, távolság stb. vonatkozásában. Itt azonban érdemes emlékezetünkbe idéznünk az ilyen típusú ítéletek státuszát Descartes érzékeléselméletében. Az érzékelés „három lépcsőfokát” Descartes a következőképpen írja le: Az első a külső tárgyak közvetlen hatása a testi érzékszervekre; a második az első közvetlen hatása az elmére, úgymint a fájdalom, a szomjúság vagy a szín. E második, ha például egy botot nézünk, „a bot által visszatükrözött színre és fényre terjed ki”; az ér-

zéki objektumot tehát nem ennek segítségével azonosítjuk az adott tárgyként (AT VII. 437). Ez utóbbi tevékenység a harmadik fázisban megy végbe, s a külső tárgyakkal kapcsolatban szokásosan meghozott ítéleteket jelenti. A harmadik lépcsőfokot Descartes csupán a közzelfogáshoz alkalmazkodva sorolja az első kettőhöz: ezen ítéletek valójában nem tartoznak az érzékeléshez, állítja, hanem kizárólag az értelemhez (vö. AT VII. 438). Az ítélet tehát nem maga az érzés, amely, mint láttuk, valamilyen módon az agyállapot „felé fordulást”, annak „szemlélését” jelenti. Ha az érzékelés során ténylegesen nem is válik el attól, az érzéki minőségek észlelése legalábbis fogalmilag nem azonos az interpretációval. A keresett mentális elem tehát nem azonos a felismerő és korrekatív ítéletekkel, és attól, hogy az elme korrekatív ítéleteket *is* hoz, még lehetne passzív. Olyan mentális elemet keresünk tehát, amely sajátosan az érzékelés második lépcsőfokához tartozik.

#### IV. ARISZTOTELIÁNUS MARADVÁNYOK DESCARTES ÉRZÉKELÉSELMÉLETÉBEN

Az érzékelés arisztotelianus-skolasztikus elképzelése vázlatosan a következő. A külső dolgok az érzékelhető minőségek formái révén színesek, rendelkeznek valamilyen ízzel stb.; ezeket a formákat, miután a megfelelő médiumon keresztül, speciestek formájában eljutottak hozzájuk, az érzékszervek felveszik vagy befogadják. Míg a tárgy „formálisan” vagy „aktuálisan” foglalja magában a minőséget, azaz például ténylegesen rózsairatú vagy piros, a közeg, például a levegő, valamint az érzékszerv nem úgy fogadja be a formát, hogy maga is rózsairatúvá vagy pirossá válik, hanem csak „intencionálisan”.

Arisztotelész esetében alapvető vita zajlik arról, hogy az érzékelés mennyiben merül ki az érzékszervek fizikai változásaiban, s a lélek szerepe nehezen állapítható meg. A középkori filozófiában – platonikus hatásra – a léleknek növekvő jelentőséget tulajdonítottak az érzékelésben. Avicennánál az érzékelés az érzékelőidegekben és az agyban zajlik, valódi szubjektuma viszont a lélek, amely a testi működéseket mintegy használja; az érzékelő aktus nem passzív befogadás, hanem fizikai változások és azok okainak aktív megragadása (Knuuttila 2008. 8–9). Az arisztotelészi *De animához* írott kommentárjában Averroës egy a cselekvő észszel analóg aktív elvvel hozza kapcsolatba az érzékelést. A lélek szerepe a továbbiakban is erősödik, és ezzel egy időben a formák vagy speciestek befogadása, ahol megjelenik, egyre bonyolultabb képet mutat. Duns Scotusnál a megismerőképeség nem pusztán befogadja tárgyai speciesteit, hanem sajátos aktussal a tárgy felé fordul (uo. 16). Nicole Oresme szerint ugyanazok a speciestek különböző érzékleteket eredményezhetnek, attól függően, hogy az érzékelőképeség hogyan interpretálja őket (Knuuttila 2008. 10–17).

A fenti megfontolások abba az irányba mutatnak, hogy az a szerep, amelyet Descartes érzékelélméletében az elmének tulajdonít, az általa elvetett hagyományból származik. Mint arra Celine Wolf-Devine – aki, mint láttuk, a mentális elemet elsősorban a korrekzív ítéletekben és a figyelem térbeli irányultságában fedezte fel – ugyancsak felfigyelt, az érzéki információ izomorf ábrákon keresztül való átadódása az érzékszerveken, az idegeken, valamint az agyon keresztül valójában a formák vagy speciestek befogadásával kapcsolatos arisztotelészi elmélet transzformációjával keletkezik (Wolf-Devine 2000. 509). A folyamat utolsó, pszichofizikai eleme véleményem szerint szintén ebben a paradigmában értelmezhető. A „szemlélés” valójában a formák megragadásának, illetve a speciestek befogadásának a maradványa. Nem beszélhetünk olyan, sajátosan karteziánus megoldásról, amelyet a Burman-beszélgetés sugallna, s amelyet a fizikai világ és az elme új felfogása, valamint a *Passions* azon ígérete alapján elvárnánk, hogy úgy tárgyalja az érzéletek és az érzelmek témáját, mintha még „soha senki nem foglalkozott volna ezzel” előtte (AT XI. 328). A szemlélésben foglalt „odafordulással” Descartes-nak nem kell aktív, akarati aktust tulajdonítania az elmének: vélhetőleg csupán annyit állít, hogy bizonyos, az agy állapotaiban manifesztálódó minőségek, s nem pedig mások azok, melyeknek az elme tudatában van a befogadás passzív módján. Az érzékelés skolasztikus magyarázata szerint e minőségeket, nem pedig a közvetítő közeg vagy az érzékszervek állapotait érzékeli a lélek<sup>4</sup> (aminek többek között abból a szempontból van jelentősége, hogy mennyiben tekinthetjük „közvetlen realista” nézeteknek a skolasztikus tanításokat).

A pszichofizikai aktus mentális részére vonatkozó descartes-i igehasználat mindig a „befogadást” sugallja. A *Traité de l'homme* számos esetben a *concevoir* igét használja:

[U]gyanis a léleknek ezen apró lökések, amelyek az agyig jutnak el az idegek közreműködésével, adnak alkalmat arra, hogy felfogja (*concevoir*) a hangok ideáját (AT XI. 149; magyarul: 20–21).

E fonalak rendeltetése ugyanis nem más, mint hogy az agyhoz közvetítés a második elemből kikerülő kicsiny részek különféle hatásait (*action*), amelyek, az eddig elmondottak alapján, alkalmat adnak a léleknek, amennyiben az egyesítve van a testtel, hogy a színek és a fény különféle ideáit ragadja meg (*concevoir*) (AT XI. 151; magyarul: 22).

Itt most nem fejtem ki, hogy sajátosan mi adhat alkalmat a léleknek arra, hogy a színek minden különbözőségét megragadja (*concevoir*), ugyanis erről fentebb eleget beszéltem (AT XI. 158; magyarul: 27).

<sup>4</sup> Ezt a skolasztikus érzékeléstan egyik legfontosabb elemeként említi Pessin 2007.

E kifejezés persze jelenthetne akármilyen nem akarati típusú mentális műveletet, hiszen Descartes *általában* a percepció, a megragadás terminusaiban beszél a gondolkodásról. A *concevoirt* azon szabály megfogalmazásánál is alkalmazza, hogy amit világosan és elkülönítetten felfogunk, az igaz: *les choses que nous concevons fort clairement et fort distinctement sont toutes vraies* („mindaz, amit igen világosan és elkülönítetten felismerünk, igaz”, AT VI. 33). Valószínűleg nem véletlen egybeesés azonban, hogy a *concevoir* latin megfelelőjét, a *concipere* igét a skolasztika és a reneszánsz számos szerzője használta az arisztotelészi értelemben vett formák, illetve speciések megragadására a *formam concipere*, illetve *speciem concipere* kifejezésekben.<sup>5</sup>

A következő szövegrészek szintén azt mutatják, hogy Descartes érzékelésemélete nem tudott kellőképpen elszakadni a formák átadásának és befogadásak paradigmájától:

Nemcsak a képzelőerő által megrajzolt képeket nevezem „ideáknak”. Sőt, amennyiben ezek a képek a testi képzeletben jelennek meg, azaz az agy valamely részén, egyáltalán nem nevezem őket „ideáknak”; csupán annyiban nevezem ideáknak őket, amennyiben *formával látják el az elmét, amikor az agy azon része felé fordul* (AT VII. 160–161. Kiemelés tőlem – Sz. J.).<sup>6</sup>

Ez történik, amikor a jó vagy a rossz a lélek közbeavatkozása nélkül megformálja önmaga lenyomatát az agyon, néha azért, mert csak a testet érinti, néha mert bár az elmét is érinti, ez nem mint jót vagy rosszat, hanem *mint valamely más formát* tekinti, amelynek az agyra kirajzolódó benyomása összekapcsolódik a jóéval vagy a rosszéval (AT XI. 398).

Nem kell meglepődnünk azon, hogy az elme részvételét az érzékelésben Descartes implicit módon a formák befogadásának módján fogja fel. Hogy a legtávolabbról indítsuk a közös jegyek említését, valójában Descartes a test–lélek egység problematikájának *ontológiai* szempontú megközelítésében is hajlamos az arisztotelészi hagyományhoz kapcsolódni. Nem egyszer beszél a lélekről vagy elméről mint a test formájáról, illetve mint olyan entitásról, amely a testnek formát ad (*informare*). (Bizonyos kommentátorok ennek alapján a test–lélek viszony hü-lomorfasta felfogását tulajdonítják Descartes-nak.)<sup>7</sup> A hajó és a kormányos között-

<sup>5</sup> *Ad decimum dicendum, quod intellectui respondet aliquid in re dupliciter. Uno modo immediate, quando videlicet intellectus concipit formam rei alicuius extra animam existentis, ut hominis vel lapidis. Thomas Aquinas 1953. Qu.1. a. 1. ad. 10. [U]t visus primo intendit in rem visibilem, secundo unitur rei visibili, tertio concipit speciem a re visibili ipsa ei admota atque praesente. Bruno 1889. 123 (Summa Terminorum Metaphysicorum, „Quando” címszó).*

<sup>6</sup> Arra vonatkozóan, hogy a „forma” kifejezést az elme ideáinak kapcsán valóban skolasztikus értelemben kell venni, lásd Pessin 2007.

<sup>7</sup> Például Rodis-Lewis 1950; Hoffman 1986.

ti viszony – amely mind ontológiai, mind érzékeléstani relevanciával bír – szintén kifejezett formában megjelenik különböző Arisztotelész-komentátoroknál (vö. Rozemond 1998. 190). Visszatérve magához a pszichofizikai folyamathoz, Descartes ugyanazt a funkciót, az élőlény fenntartásának funkcióját tulajdonítja az érzékelésnek, mint Arisztotelész (vö. Lautner 2006. 206). A cselekvésnek és a szenvedésnek csupán a test, illetve a lélek vonatkozásában fennálló különbsége az érzékelés kontextusában, mint említettem, szintén arisztotelészi eredetű. A szemantikai felfogás, amelyet bizonyos értelmezők Descartes-nál felismertek – és amelyről megállapítottuk, hogy felhasználása Descartes értelmezésében bizonyos tekintetben megvilágító erejű, ám nem szolgál a descartes-i interakció-felfogás valódi magyarázatául – ugyancsak skolasztikus szerzőkhöz kapcsolható (lásd erről Behan 2000). Az érzékelés folyamatának leírása pedig, mint említettem, a formák befogadásának mintegy a mechanisztikus fiziológiába áthangolt változatát nyújtja.

Descartes számos alkalommal hangsúlyozza érzékelésfelfogásának antiarisztotelianus, antiskolasztikus jellegét. Ha azonban közelebről megnézzük az arisztotelianus érzékeléstan Descartes által kritizált elemeinek körét, valamint a kritika jellegét és megalapozottságát, arra a következtetésre jutunk, hogy szerzőnk elsősorban egy bizonyos tézis cáfolatára törekedett, és ellenvetései nem tanúskodnak az arisztotelianus megoldási kísérletek alapos ismeretéről.

A tézis, melyet Descartes elvetett, leegyszerűsítve úgy szól, hogy az érzékelés során az érzékelt tárgyból annak valamiféle hasonmása „jut be” az érzékelő szubjektumba, s ily módon az érzéketek tartalma hasonlít a külső dologhoz, mely létrehozta. Az érzékelés azon aspektusai, melyekkel kapcsolatban Descartesnak vitája van az arisztotelianusokkal, nem az elme hozzájárulásának módjára, hanem az érzéketek tartalmára, illetve a fiziológiai történések jellegére vonatkoznak.

Descartes elsődleges és legátfogóbb filozófiai célja minden bizonnyal a fizikai világ új, antiarisztotelianus leírása volt, amely az emberi test önálló működését is magában foglalja. Egészen más jellegű ambíciója Descartes-nak a lélek potenciális elválaszthatóságának bizonyítása a *Meditationes*-ban, de egyik is, másik is a két szubsztancia önállóságának belátása irányában hat. Descartes a test–lélek egység problematikájának és a pszichofizikai működéseknek mint olyanoknak közismerten kevés figyelmet szentelt. Az érzékelés *fiziológiai* leírása igen aprólékos; a lélek szerepének magyarázatára azonban gyakorlatilag csak ígéreteket kapunk (például a *Traité de l'homme*-ban és az Erzsébet-levelekben). A *Passions*, amely a legrészletesebben tárgyalja e problémákat, Erzsébet hercegnő kérdéseinek nyomására, sietve keletkezett (lásd AT XI. 326). Érthető tehát, hogy Descartes hanyagul és feltehetően reflektálatlanul átvette az arisztotelészi-skolasztikus hagyomány felfogását a lélek szerepéről a pszichofizikai aktusban, különösen mert ez olyan elméletből származván, amellyel szemben a test–lélek probléma jóval kevésbé élesen merül fel, némiképp az utóbbi észrevétlen elkerülésének



lehetőségével is kecsgetethetett. Ahogyan Descartes a *Dioptrique*-ban írja, érzékelélmélete megpróbált „a lehető legkevésbé eltérni a bevett nézetektől” (AT VI. 112). Véleményem szerint jobban sikerült megfelelnie e célkitűzésnek, mint általában véljük.

## IRODALOM

- Aho, Tuomo 2007. Suárez on Cognitive Intentions. In Paul Bakker – Johannes Thijssen (szerk.) *Mind, Cognition and Representation: The Commentary Tradition of Aristotle's De Anima*. Aldershot, Ashgate. 179–204.
- Arisztotelész 2006. *Lélekfilozófiai írások*. Steiger Kornél fordítását átdolgozta Brunner Ákos és Bodnár István. Budapest, Akadémiai Kiadó.
- Behan, David 2000. Descartes and Formal Signs. In Stephen Gaukroger – John Schuster – John Sutton (szerk.) *Descartes' Natural Philosophy*. London – New York, Routledge. 528–541.
- Bruno, Giordano 1889. *Iordani Bruni Nolani Opera latine conscripta publicis sumptibus edita*. Vol. I, pars IV. Szerk. Felice Tocco – Girolamo Vitelli. Florentiae.
- Descartes, René 1996. *Œuvres de Descartes*. Szerk. Charles Adam – Paul Tannery. Paris, Vrin (= AT).
- Descartes, René: *Értekezés az emberről*. Gulyás Péter fordítása (kézirat).
- Gaukroger, Stephen – John Schuster – John Sutton (szerk.) 2000. *Descartes' Natural Philosophy*. London – New York, Routledge.
- Hoffman, Paul 1986. The Unity of Descartes' Man. *Philosophical Review*. 95. 339–370.
- Knuutila, Simo 2008. Aristotle's Theory of Perception and Medieval Aristotelianism. In Simo Knuutila – Pekka Kärkkäinen (szerk.) *Theories of Perception in Medieval and Early Modern Philosophy*. Dordrecht, Springer. 1–22.
- Lautner Péter 2006. Arisztotelész lélekfilozófiája a legújabb kutatások tükrében. In Arisztotelész 2006. 193–218.
- Loeb, Louis 1981. *From Descartes to Hume: Continental Metaphysics and the Development of Modern Philosophy*. Ithaca–London, Cornell University Press.
- Nadler, Steven 1997. Critical Notice of Gordon Baker and Katherine Morris, *Descartes' Dualism*. *Philosophical Books*. 38. 157–164.
- Pessin, Andrew 2007. Descartes' Theory of Ideas. In Edward N. Zalta (szerk.) *The Stanford Encyclopedia of Philosophy* (Spring 2007 Edition).  
URL: <http://plato.stanford.edu/entries/descartes-ideas/> Hozzáférés: 2010. 10. 01.
- Rodis-Lewis, Geneviève 1950. *L'individualité selon Descartes*. Paris, Vrin.
- Rozemond, Marleen 1998. *Descartes' Dualism*. Cambridge/MA, Harvard University Press.
- Stump, Eleonore 2004. *Aquinas*. London – New York, Routledge.
- Thomas Aquinas 1953. *Quaestiones Disputatae de Potentia*. Taurinum, Marietti.
- Wilson, Catherine 1993. Constancy, Emergence, and Illusions: Obstacles to a Naturalistic Theory of Vision. In Steven Nadler (szerk.) *Causation in Early Modern Philosophy: Cartesianism, Occasionalism, Preestablished Harmony*. Penn State University Press. 159–178.
- Wolf-Devine, Celia 2000. Descartes' Theory of Visual Spatial Perception. In Stephen Gaukroger – J. Schuster – J. Sutton (szerk.) *Descartes' Natural Philosophy*. London – New York, Routledge. 506–523.
- Yolton, John W. 1996. *Perception and Reality: A History from Descartes to Kant*. Ithaca–London, Cornell University Press.

E. SZABÓ LÁSZLÓ – GYENIS BALÁZS –  
GYENIS ZALÁN – RÉDEI MIKLÓS – SZABÓ GÁBOR

## Korrelációk kauzális magyarázata

*Nincs korreláció kauzalitás nélkül* – így hangzik tömören az a metafizikai elv, amelyre a tudományfilozófiában *Reichenbach közös ok elve*ként szokás hivatkozni. Kevésbé tömören fogalmazva a közös ok elve azt állítja, hogy a világban felmutatható bármely korreláció vagy a korreláló események közötti közvetlen kauzális kapcsolatból eredeztethető, vagy egy harmadik eseményre, a korreláló események úgynevezett közös okára vezethető vissza. Az alábbiakban arról a mintegy tizenöt évet átfogó munkáról szeretnénk rövid áttekintést nyújtani, amelyet e tanulmány szerzői a reichenbachi közös ok elv, vagyis a korrelációk kauzális magyarázatának tanulmányozása terén folytattak.

### BEVEZETÉS

A reichenbachi közös ok elv szerint bármely két olyan korreláló eseménynek, amelyek nem állnak egymással közvetlen kauzális vagy logikai kapcsolatban, létezik közös oka. Habár az elv nevében Hans Reichenbachra (1956) vezethető vissza, aki a közös ok fogalmát először öntötte explicit formába, az elv mégsem tőle, hanem tanítványától, Wesley Salmontól (1978) ered. Salmon az elvet abban a vitában fogalmazta meg először, amelyet egy jó évtizeden keresztül Bas C. van Fraassenel (1982) folytatott a tudományos magyarázat mibenlétéről. Salmonnak a vitában azért volt szüksége a közös ok elvre, mert ennek az elvnek a segítségével vélte elkülöníthetőnek egymástól a valódi tudományos magyarázat alapjául szolgáló kauzális folyamatokat az úgynevezett pszeudofolyamatoktól, mint amilyen például egy árnyék terjedése a falon. A pszeudofolyamatok olyan korreláló eseményekből állnak, amelyek között nincs közvetlen kauzális viszony, hanem a korrelációt egy közös ok hozza létre.

Ettől a tudományfilozófiai problémától függetlenül a hetvenes évektől kezdődően egy másik tudományterületen is egyre több figyelmet kapott a reichenbachi közös ok elv, nevezetesen a kvantumelmélet rejtett paraméteres kutatásaiban. Atomi objektumokon végzett mérések kvantumelméleti jóslatai, valamint

később a ténylegesen elvégzett mérések ugyanis azt sugallták, hogy léteznek olyan, távoli események közötti korrelációk, amelyek esetében kizárható mind a közvetlen kauzális hatás, mind a közös ok. Röviden, a reichenbachi közös ok elv a kvantumjelenségek körében nem érvényes – hangzott a verdikt.

Így aztán egyre inkább a körül a kérdés körül kezdett el forogni a vita, hogy miféle elv is a reichenbachi közös ok elv. Metafizikai elv? Empirikus állítás? Metodológiai heurisztika? Ha pedig empirikus állítás vagy metafizikai elv, akkor hol húzódnak érvényességének határai? Érvényes a klasszikus fizikában, de nem érvényes a kvantumelméletben? Kiterjeszthető az elv a kvantumtérelméltre, vagy sérül a véges jelterjedési sebesség következményeként?

Ebben a tanulmányban nincs módunk a kérdésre adott különféle válaszokat áttekinteni.<sup>1</sup> Annyit azonban leszögezhetünk, hogy a nyolevanas évek végétől az irodalomban az a konszenzus kezdett kibontakozni, hogy a reichenbachi közös ok elv *nem érvényes* univerzálisan. Ezt az álláspontot olyan kvantumelméleti, illetve klasszikus fizikai szituáció felmutatásával igyekeztek alátámasztani, amelyekben adva van egy korreláció anélkül, hogy a korreláló események közvetlen kauzális kapcsolatban állnának, vagy közös okkal rendelkeznének (Sober 1988; Van Fraassen 1989).

Mindezen cáfolatkísérleteknek közös problémája volt azonban, hogy a közös ok elv érvényességét fizikai szituációk egyfajta informális megközelítésén keresztül, intuitív érvelésre hagyatkozva próbálták eldönteni. Könnyű azonban belátni, hogy mind a fizikai szituációkat, mind a közös ok elvét sokféleképpen lehet rekonstruálni. Mit tekintünk egy adott fizikai szituációban eseményeknek? Mit tudunk ezen események közvetlen oksági viszonyairól? Mikor mondjuk, hogy két esemény korrelál egymással? Milyen feltételek mellett tekinthetünk egy harmadik eseményt két korreláló esemény közös okának? Nem meglepő módon különféle rekonstrukciók esetén különféle válaszokat kaphatunk arra a kérdésre, hogy két korreláló, egymással oksági viszonyban nem lévő eseménynek vajon van-e közös oka. Ezért a közös ok elvének érvényességét nem lehet azelőtt eldönteni, hogy a feladat által megkívánt módszerességgel rögzítettük volna, mit is értünk az elvben szereplő fogalmak alatt. Mivel semmi sem zárja ki, hogy a fogalmaknak több, valamilyen szempontból természetes megfogalmazása is létezen, törekedni kell arra, hogy a lehetőségek szerint minél általánosabb keretek között tárgyaljuk a problémát.

Az alábbiakban azt a kutatási programot szeretnénk röviden bemutatni, amely a reichenbachi közös ok elvében szereplő fogalmak pontos valószínűségelméleti megfogalmazására épül.<sup>2</sup>

<sup>1</sup> Ehhez lásd Szabó G. 2006 (a szerző idegen nyelven Gábor Hofer-Szabó néven publikál).

<sup>2</sup> Jelen tanulmány nem érinti a kauzális gráfelméleti rekonstrukció eredményeit. Ehhez a kutatási irányhoz lásd pl. Pearl (2000) és Spirtes (2000) munkáit.

## A REICHENBACHI KÖZÖS OK

A reichenbachi közös ok elv illusztrációját kezdjük egy egyszerű példával. Mi a magyarázata annak, hogy a futballszurkolók a meccsen többnyire egyszerre ugranak fel; röviden, hogy a felugrások korrelálnak? A közös ok elve azt mondja ki, hogy ez a korreláció vagy a szurkolók közötti közvetlen kauzális kapcsolatból ered, vagy valamilyen harmadik eseményből. Kis tanmesénk mindkét esetre példaként szolgálhat. A korreláció eredhet közvetlen kauzális kapcsolatból is, ha mondjuk a szurkolók egyszerűen azért ugranak fel, mert nem látnak az előttük felállóktól; vagy származhat egy közös okból is: a pályán kialakult, a szurkolókat lázba hozó helyzetekből.

Milyen matematikai fogalmak révén ragadható meg a fenti szituáció? Először a korreláció fogalmát kell tisztáznunk. Két szurkoló felugrásai közötti (pozitív) korrelációt azt jelenti, hogy két szurkoló többször ugrik fel egyszerre, mint azt a külön-külön felugrások alapján várnánk. Tekintsük a meccset mondjuk egyperces felbontásban, és tegyük fel, hogy két kiválasztott szurkolónk tízszer ugrott fel a meccs alatt, mind a tízszer ugyanabban a percben. Ekkor az együttes felugrások aránya a 90 perchez  $10/90$ , a külön-külön felugrások arányának szorzata viszont csak  $10/90 \times 10/90$ , vagyis kilencszer kisebb. A felugrások tehát korrelálnak.

Forduljunk most a bennünket érdeklő közös ok típusú kauzális magyarázat felé, és tegyük fel, hogy mind a tíz felugrás azért történt, mert a pályán valamilyen izgalmas helyzet alakult ki. Az egyszerűség kedvéért tegyük fel továbbá azt is, hogy a kauzális viszony determinisztikus: minden helyzet felugrást eredményezett, és csak az ilyen helyzetekben történt felugrás. Ekkor a következőket mondhatjuk. (1) Ha a 90-ből csak azokat a perceket tekintjük, amikor a pályán helyzet állt elő, akkor a felugrások közötti korreláció *eltűnik*, ugyanis ebben a tíz esetben mind a külön-külön felugrások, mind az együttes felugrások aránya a szóban forgó 10 perchez viszonyítva 1, hiszen minden ilyen percben a külön-külön és (így) az együttes felugrások is bekövetkeztek. (2) Ha a 90-ből csak azokat a perceket tekintjük, amikor a pályán nem volt helyzet, akkor a felugrások közötti korreláció *szintén eltűnik*, ugyanis ebben a 80 esetben mind a külön-külön felugrások, mind az együttes felugrások aránya a szóban forgó 80 perchez 0 (nem következett be sem külön-külön, és így az együtt felugrás sem). (3) Az, hogy *A* szurkoló felugrott, többször következett be (valójában mindig bekövetkezett) akkor, amikor helyzet volt a pályán, mint akkor, amikor nem. (4) Hasonlóan, az, hogy *B* szurkoló felugrott, többször következett be akkor, amikor helyzet volt, mint akkor, amikor nem.

A közös ok és a korreláció viszonyát a fenti négy kritérium tehát a következőképpen jellemzi. (1) és (2) azt állítja, hogy a két esemény közötti korreláció minden olyan statisztikus mintán eltűnik, amelyben a közös ok egyértelműen fennáll, vagy amelyben egyértelműen nem áll fenn. Ezt a jelenséget nevezzük

*árményekolásnak (screening off)*. A (3) és (4) kritérium pedig azt állítja, hogy a közös ok bekövetkezése – intuitíve szólva – „növeli” mind az egyik, mind a másik esemény bekövetkezésének esélyét. E négy kritériumot tekintette Reichenbach a közös ok karakterizációjának: egy korreláció közös oka csak olyan esemény lehet, amely legalább a fenti kritériumoknak megfelel.

Ez a karakterizáció akkor kap különös hangsúlyt, ha a determinisztikus esetről áttérünk az indeterminisztikus vagy másképp sztochasztikus esetekre. Tegyük fel, hogy a szurkolók nem minden esetben ugranak fel egyszerre, de azért felugrásaik korrelálnak. Mi a magyarázata a korrelációnak? Nyilván a pályán kialakult helyzetek. A felugrások és a helyzetek közötti viszony azonban nem determinisztikus, vagyis a szurkolók néha nem ugranak fel egy-egy helyzet láttán, vagy felugranak akkor is, ha a pályán nem történik semmi. Milyen értelemben közös okai tehát a pályán történtek a korrelációnak? A válasz a fenti kritériumok valószínűségi általánosításában rejlik. A pályán történtek valószínűségi értelemben a korreláció közös okának tekinthetők, ha a fenti négy kritérium általánosítása fennáll a közös okra és a korrelációra nézve; vagyis a felugrásokat mind a helyzetek, mind azok hiánya leárnyékolja ([1]–[2] kritérium), illetve a helyzetek növelik a felugrások valószínűségét mindkét szurkoló esetében ([3]–[4] kritérium). Ahhoz azonban, hogy a kritériumokat ebben a sztochasztikus esetben is pontosan meg tudjuk fogalmazni, be kell vezetnünk a valószínűség fogalmát.

Egy jelenségkör leírásához használt valószínűségi modell két alapvető komponensből áll: egy eseményalgebrából és a rajta értelmezett valószínűségből. „Algebráról” itt abban az értelemben beszélünk, hogy tetszőleges  $A$  és  $B$  eseményre értelmezettnek gondoljuk az „ $A$  és  $B$ ”, az „ $A$  vagy  $B$ ” illetve a „nem  $A$ ” eseményeket. Egy ilyen valószínűségi modell szokásos matematikai reprezentációja egy úgynevezett valószínűségi mértéktérrel történik, amely alatt egy  $(X, \Sigma, p)$  hármast értünk, ahol  $X$  egy halmaz,  $\Sigma$  az  $X$  halmaz részhalmazaiából képzett  $\sigma$ -algebra, vagyis egy olyan struktúra, amely zárt a részhalmazok közötti bizonyos halmazelméleti műveletekre (megszámlálható unió és metszet, valamint komplementáció; a megszámlálhatóan végtelen unióra és metszetre való zártságot jelöli az algebra előtt a  $\sigma$ ),  $p$  pedig egy  $\sigma$ -additív, normált mérték, azaz egy olyan leképezés  $\Sigma$  elemeiről a  $[0, 1]$  valós számokra, amelyre teljesülnek az alábbi követelmények:

$$i. \quad p(X) = 1$$

$$ii. \quad p\left(\bigcup_{i=1}^{\tau} A_i\right) = \sum_{i=1}^{\tau} p(A_i), \text{ ahol az } A_i \in \Sigma \text{ halmazok páronként diszjunktak.}$$

Nyilvánvalóan a mértékelméleti megfogalmazásban  $\Sigma$  tölti be az eseményalgebra szerepét,  $p$  pedig a valószínűség.

Klasszikus valószínűségi mértéktérre a paradigmaticus példa a szabályos dobókockát reprezentáló mértéktér. Itt  $X$  az  $\{1,2,3,4,5,6\}$  halmaz,  $\Sigma$  az  $X$  részhalmazából képzett 64 elemű halmaz, amely olyan elemeket tartalmaz, mint a  $\{2,4,6\}$  páros dobás, az  $\{1,2,3,4\}$  ötnél kisebb dobás, vagy mondjuk a  $\{6\}$  hatos dobás. A  $p$  mérték pedig az az  $i$ -ii. tulajdonságokat kielégítő  $p: \Sigma \rightarrow [0,1]$  hozzárendelés, amely mindegyik  $\{i\}$  elemhez ( $i = 1 \dots 6$ )  $1/6$ -ot rendel. Vegyük észre, hogy ezzel  $p$ -t az egész  $\Sigma$  eseményalgebrán megadtuk. Például a páros dobás  $1/2$  valószínűségű lesz, az ötnél kisebb dobás  $2/3$  valószínűségű stb.

A továbbiakban szükségünk lesz még egy fogalomra, a feltételes valószínűség fogalmára. Egy  $A$  eseménynek egy  $B$  eseményre vett  $p(A|B)$  *feltételes valószínűségét*, amennyiben  $p(B) \neq 0$ , az alábbi összefüggéssel definiáljuk:

$$p(A|B) = \frac{p(A \cap B)}{p(B)},$$

ahol  $A \cap B$  az  $A$  és  $B$  események metszete.

Arra a kérdésre, hogy mit is *jelent* az, hogy a hatos dobás valószínűsége  $1/6$ , a fenti matematikai modell nyilvánvalóan nem ad választ. A valószínűségnek többfajta interpretációja létezik, itt azonban az interpretációk kérdésében nincs módunk elmerülni. Az egyszerűség kedvéért valószínűség alatt értsünk relatív gyakoriságot: a hatos dobás valószínűsége akkor  $1/6$ , ha dobások egy elegendően hosszú sorozatában a dobások közel egy hatoda hatos. Hogy mi számít elegendően hosszú sorozatnak, vagy mit jelent a „közel” kitétel, azzal most nem foglalkozunk.

Ha adva van egy klasszikus valószínűségi mértéktér, akkor két esemény korrelációját könnyen megfogalmazhatjuk. Egy  $A$  és  $B$  esemény akkor *korrelál pozitívan* a  $(X, \Sigma, p)$  mértéktéren, ha fennáll a következő egyenlőtlenség:

$$p(A \cap B) > p(A) \cdot p(B),$$

vagyis ha az együttes esemény valószínűsége nagyobb, mint az egyes események valószínűségeinek szorzata. Ha az egyenlőtlenséggel fordítva áll, akkor *negatív korrelációról*, ha pedig egyenlőtlenség helyett egyenlőség áll, akkor *valószínűségi függetlenségről* beszélünk. Fontos hangsúlyozni, hogy mind a korreláció, mind a függetlenség csak azután értelmes fogalom, hogy rögzítettük a valószínűségi mértéktérre.

És most következzen a reichenbach-i közös ok definíciója. Legyen  $A$  és  $B$  két pozitívan korreláló esemény az  $(X, \Sigma, p)$  mértéktéren. Egy harmadik  $C$  eseményt ugyanebben a mértéktérben akkor nevezünk az  $A$  és  $B$  esemény közötti korreláció *közös okának*, ha a fenti négy kritériumot kielégíti, vagyis ha a közös ok (1) mind jelenlétével, (2) mind távollétével függetlenné teszi az  $A$  és  $B$  eseményt, továbbá jelenlétével növeli (3) mind az  $A$  esemény, (4) mind a  $B$  esemény valószínűségét. Ezek a feltételek a feltételes valószínűség segítségével könnyen megfogalmazhatók, és az alábbi alakot öltik:

$$p(A \cap B | C) = p(A | C) \cdot p(B | C), \quad (1)$$

$$p(A \cap B | \neg C) = p(A | \neg C) \cdot p(B | \neg C), \quad (2)$$

$$p(A | C) > p(A | \neg C), \quad (3)$$

$$p(B | C) > p(B | \neg C). \quad (4)$$

A fenti négy összefüggés Reichenbach híres közös ok-definíciója, amely egyben a közös ok első valószínűségi megfogalmazása is. A definíció Reichenbach *The Direction of Time* (1956) című művéből származik, amelyben Reichenbach a közös okot az idő aszimmetriájának, a múlt és a jövő különbségének megalapozására használta – több filozófus (például Huw Price 1996. 118) szerint nem kellően megalapozottan. Hogy a reichenbach-i kritériumok mennyiben ragadják meg helyesen a közös ok fogalmát, természetesen sokat vitatott kérdés. A definíció a közös oknak nyilvánvalóan legfeljebb a szükséges feltételeit adhatja meg, hiszen a  $C$  esemény minden determinisztikus okozata, amely egyazon statisztikát fog követni, mint  $C$ , ki fogja elégtíteni a fenti kritériumokat anélkül, hogy maga az  $A$  és  $B$  közötti korreláció közös oka lenne. Az  $A = C$  vagy  $B = C$  esemény szintén kielégíti a definíciót, holott a korreláló eseményeket magukat nyilván nem fogadjuk el közös okként. Ugyanakkor a kritériumok szükséges voltát is sokan kétségbe vonták, vagy a definíciónak egy általánosabb formáját tekintették a közös ok helyes karakterizációjának. Minderre a tanulmány második felében még visszatérünk. Most azonban fogadjuk el, hogy a reichenbach-i kritériumok a közös okra vonatkozó intuíciónk helyes valószínűségelméleti megfogalmazásai. Mi következik ebből a reichenbach-i közös ok elvére nézve?

## KIBŐVÍTHETŐSÉG

A közös ok reichenbach-i definíciója nyilvánvalóan feltételezi, hogy  $A$ ,  $B$  és  $C$  események ugyanahhoz az eseménytérhez tartoznak: ha ez nem így volna, az (1)–(4) összefüggéseknek nem lenne értelme. Ez más szóval azt jelenti, hogy ha a reichenbach-i közös ok elv teljesül, akkor az események egy kellően teljes leírását nyújtó valószínűségi elmélet eseményalgebrájának a korreláló eseménypárok mellett a közös okokat is tartalmaznia kell. Az elv tehát hallgatólagosan feltételezi, hogy a szóban forgó jelenségeket leíró valószínűségi elméletünk eseményalgebrája szükség esetén mindig kiterjeszthető olyanná, hogy az tartalmazza a közös okokat reprezentáló elemeket a szükséges valószínűségi tulajdonságokkal. Ez a kiterjesztés azt a szituációt reprezentálja, amikor a kérdéses korrelációnak „rejtett” közös okai vannak – „rejtettek” abban az értelemben, hogy nem jelennek meg a szituációt durván modellező eredeti  $(X, \Sigma, p)$  mértéktérben, viszont jelen lennének akkor, ha a jelenségekről finomabb leírást adnánk egy kibővített  $(X', \Sigma', p')$  mértéktér segítségével.

Felmerül a kérdés, hogy egy ilyen kiterjesztés mindig lehetséges-e. Érdekes módon ezt a kérdést sokáig senki sem vetette fel. Mint vizsgálatainkból kiderült, a probléma egyáltalán nem triviális.

A kérdés megválaszolása azonban pontos előkészületeket igényel. Mindenekelőtt definiálni kell, hogy mit értünk egy valószínűségi mértéktér konzisztens kiterjesztésén, vagyis olyan kiterjesztésén, amely az eredeti események algebrai és mértékelméleti tulajdonságait érintetlenül hagyja, ugyanakkor újabb eseményeket illeszt az eseménytérbe.<sup>3</sup> Csak ezek után vethető fel a kérdés, hogy egy algebra lehetséges konzisztens kiterjesztései között van-e olyan, amely tartalmazza a szűkebb algebra valamely korrelációjának közös okát. Klasszikus valószínűségi mértéktér esetében a kérdésre a válasz a következő: Az eredeti mértéktér korrelációinak *tetszőleges* véges halmazához létezik a mértéktér olyan kiterjesztése, hogy a kiterjesztett mértéktérben a halmazba tartozó korrelációk mindegyikének van közös oka (Hofer-Szabó – Rédei – E. Szabó 1999).

Ha egy algebra több korreláló eseménypárt tartalmaz, akkor a kiterjesztési procedura ismétlésével az összes korrelációhoz közös ok található. Az eljárás a következő. Kiválasztunk egy  $(A_1, B_1)$  korreláló párt az eredeti  $(X, \Sigma, p)$  mértéktérben, majd pedig kiterjesztjük az  $(X, \Sigma, p)$  mértéktérrel egy  $(X', \Sigma', p')$  mértéktérre, amely már tartalmazza a korreláció  $C_1$  közös okát. Ezek után veszünk egy másik  $(A_2, B_2)$  korreláló párt az eredeti  $(X, \Sigma, p)$  mértéktérben. Mivel a kiterjesztés konzisztens, ezért ez a korreláció reprezentálva lesz a bővebb  $(X', \Sigma', p')$  mértéktérben is, ezért az eljárás folytathatjuk: kiterjesztjük az  $(X', \Sigma', p')$  mértéktérrel egy  $(X'', \Sigma'', p'')$  mértéktérre, amely már tartalmazza az  $(A_2, B_2)$  korreláció  $C_2$  közös okát – és így tovább. Az egymást követő kiterjesztésekkel tehát olyan valószínűségi elméletet nyerünk, amely az eredeti elmélet tetszőleges véges számú korrelációjára nézve tartalmaz közös okot. Valójában ennél még több is megmutatható: az egymást követő kiterjesztések során a közös okok tetszőleges, a reichenbach-i (1)–(4) feltételeket kielégítő valószínűségi tulajdonságokkal rendelkezhetnek (a részleteket lásd Hofer-Szabó – Rédei – E. Szabó 1999).

## A KÖZÖSOK-ZÁRTSÁG

Első ránézésre ezeket az eredményeket úgy is értékelhetnénk, hogy bármilyen legyen is a világ kauzális rendje, és bármilyen legyen is a világ eseményeinek korrelációs struktúrája, nincs annak valószínűségelméleti akadálya, hogy a kettő a Reichenbach-féle közös ok elvnek megfelelő összhangban legyen. A helyzet

<sup>3</sup> Pontosabban egy  $(X', \Sigma', p')$  valószínűségi mértéktér akkor konzisztens kiterjesztése egy  $(X, \Sigma, p)$  mértéktérnek, ha  $\Sigma$ -nak létezik egy olyan  $h$  Boole- $\sigma$ -algebra-beágyazása (az algebrai műveleteket megőrző, az elemeket nem „összecséjtő” leképezése)  $\Sigma'$ -be, hogy  $\Sigma$  minden  $A$  elemére fennáll a következő összefüggés:  $p'(h(A)) = p(A)$ .



azonban nem ennyire egyszerű. Ha ugyanis feltesszük, hogy az események korrelációs struktúrája leírható (klasszikus) valószínűségi eszközökkel, akkor a közös ok elv csak úgy lehet igaz, ha azt is feltesszük, hogy (elvből) létezik olyan  $(X, \Sigma, p)$  valószínűségi mértéktér, hogy a  $\Sigma$  eseményalgebrában található összes olyan korreláló eseménypárhoz, amely nincs direkt kauzális kapcsolatban egymással – jelölje  $R_{ind}$  ezek halmazát – található legyen közös ok *magában a  $\Sigma$  eseményalgebrában*. Hangsúlyoznunk kell, hogy a kiterjeszhetőségi tételekből ez nem feltétlenül következik, hiszen itt csak azt sikerült bizonyítanunk, hogy tetszőleges véges számú korrelációnak megadható a közös oka a kiterjesztett elméletben<sup>4</sup> – az  $R_{ind}$  halmaz viszont lehet *végtelen*. Vegyük észre, hogy az sem oldja meg a helyzetet, ha feltételezzük, hogy az elméletünkben eredetileg csak véges sok eseménypár közötti korreláció vár közös ok típusú magyarázatra. A kiterjesztések ugyanis új eseményeket hoznak be, és ezáltal új korrelációk tűnhetnek fel. Vagyis a kiterjesztések révén egyfelől képesek leszünk közös okkal megmagyarázni az eredeti algebrában magyarázatra szoruló összes (véges sok) korrelációt, másfelől viszont új, (szintén véges sok) esetleg  $R_{ind}$ -be tartozó, tehát magyarázatra szoruló korrelációt involválunk.

Ezzel összefüggésben felmerülő érdekes kérdés, hogy léteznek-e olyan valószínűségi mértékterek, melyekben *minden* korrelációhoz létezik a közös októl elvárt reichenbach-i feltételeket kielégítő esemény. Az ilyen valószínűségi mértéktereket *közösok-zárt mértéktereknek* nevezzük. A válaszhoz először is meg kell különböztetnünk a valószínűségi értelemben atomos és a nem atomos mértéktereket. Egy mértéktérben valószínűségi atomnak nevezzük a legkisebb nem nulla valószínűségű elemeket. A közösok-zártság tekintetében az atomos és nem atomos mértékterek homlokegyenest ellenkező képet mutatnak. Az atomos mértékterek (a triviális, egyatomos mértéktér kivételével) nem közösok-zártak, a nem atomos mértékterek viszont kivétel nélkül azok. De ami még fontosabb, *bármely* nem közösok-zárt klasszikus valószínűségi mértéktér konzisztens módon kiterjeszhető, vagyis beágyazható egy olyan mértéktérbe, amely már közösok-zárt (Gyenis B. – Rédei 2004; valamint Gyenis Z. – Rédei 2010).

Természetesen nem minden kauzális viszony közös ok típusú. Így a fenti tételek túl erősek ahhoz, hogy igazán jelentős következtetéseket vonhassunk le belőlük a korrelációk kauzális magyarázatára vonatkozóan. A valószínűségi struktúra önmagában nem mond semmit az események időbeli viszonyáról, így például a korrelációk nemesak közös okból eredhetnek, hanem közvetlen kauzális hatásból is. A közösok-zártság vizsgálatához gyümölcsözőbb kiindulási pontot jelent tehát az, ha először rögzítjük a korrelációknak azon osztályát, amelyre *nem* kívánunk közös ok típusú magyarázatot adni, mégpedig azért, mert a szó-

<sup>4</sup> Érdekes módon éppen a bonyolultabb kvantum-valószínűségi modellekben sikerült ennél erősebb tételt bizonyítanunk, korrelációk tetszőleges halmazára nézve (Hofer-Szabó – Rédei – E. Szabó 1999).

ban forgó korreláló események között közvetlen kauzális viszony van. Ilyenkor formálisan úgy célszerű eljárunk, hogy előre rögzítünk egy  $R_{ind}$  relációt, amely a kauzálisan független események között áll fenn, és a közösok-zártsággal kapcsolatos vizsgálódásainkat az ilyen relációban álló eseményekre korlátozzuk. Az  $R_{ind}$  reláció hangolásával ezek után skalázhatjuk a különféle kauzális szituációkat attól függően, hogy milyen erős közvetlen kauzális viszony mellett keressük a maradék korrelációk közös ok típusú magyarázatát. A témában még sok nyitott kérdés van, annyi azonban már most megállapítható, hogy, bár a véges valószínűségi mértékterek kauzális zártsága lehetséges még elegendően gyenge és észszerű  $R_{ind}$  reláció mellett is, a kauzális zártság tipikusan nem-véges valószínűségi mértékteret igényel.

### A KÖZÖS OK ÉS A KÖZÖS KÖZÖS OK

Láttuk, hogy a reichenbachi közös ok elve nem cáfolható azon a módon, hogy felmutatunk egy nem közösok-zárt valószínűségi mértékteret, mivel minden klaszikus valószínűségi mértéktér konzisztensen kiterjeszhető olyan mértéktérre, amely már tartalmazza az eredeti mértéktér korrelációinak közös okát. Szintén láttuk, hogy a kiterjesztés lépésként haladt: első lépésben az  $(A_1, B_1)$  korrelációhoz kerestünk egy közös okot, majd egy következő lépésben a  $(A_2, B_2)$  korrelációhoz, és így tovább. Fel kell azonban hívnunk a figyelmet arra, hogy a második lépésben talált  $C_2$  közös ok nem feltétlenül egyezik meg az első lépésben talált  $C_1$  közös okkal, sőt általában ez a helyzet – a két korrelációnak különböző közös okai lesznek. Ez tökéletesen megfelel hétköznapi intuíciónknak: miért is lenne két különböző jelenségnek ugyanaz a kauzális magyarázata, miért is rendelkezne két korreláció egyetlen úgynevezett *közös* közös okkal?

Érdeemes röviden megvilágítanunk a közös ok *vs.* *közös* közös ok megkülönböztetés hátterét. A két fogalom különbségére először Nuel Belnap hívta fel a figyelmet, és 1996-ban történik róla először említés az irodalomban (Belnap – E. Szabó 1996). A különbségtétel a következő fejezetben ismertetésre kerülő közösok-rendszerek kontextusában merült fel, melynek az eddig tárgyalt reichenbachi közösok-fogalom egy speciális esete. A közösok-rendszer fogalma lényegében megegyezik azzal, amit az EPR–Bell-probléma irodalmában a fizikusok rejtett paraméternek neveznek. Az EPR–Bell-problémára még visszatérünk. Lényege, hogy bizonyos kvantummechanikai kísérletekben olyan korrelációkat figyelhetünk meg, amelyeknek – a szokásos argumentumok szerint – nem létezhet sem direkt, sem közös ok típusú magyarázata. Nevezetesen, ha feltesszük valamilyen rejtett paraméter létezését, akkor e feltételezésből olyan egyenlőtlenségek vezethetők le (Bell-egyenlőtlenségek), melyeket a kísérletben megfigyelt valószínűségek sértenek. Ebben a kontextusban fontos volt észrevenni, hogy a Bell-egyenlőtlenségek levezetésében hallgatólagosan feltesszük, hogy a

különböző korrelációknak *közös* közös oka van (pontosabban *közös* közösok-rendszer, lásd alább) egy közös rejtett paraméter formájában.

Míndezek tükrében a Reichenbach-féle közös ok kontextusában is megkérdezhetjük, vajon nem igaz-e az is, hogy a klasszikus valószínűségi terek kiterjeszthetők úgy, hogy a korreláló eseménypároknak a kiterjesztett mértékterben *ugyanaz* az esemény legyen a közös oka. Vagyis található-e a korrelációk tetszőleges halmazához egy *közös* közös ok? A válasz nemleges (Hofer-Szabó – Rédei – E. Szabó 2002): a klasszikus valószínűségi mértékterek általában nem terjeszthetők ki úgy, hogy bennük több különböző korrelációnak ugyanaz legyen a közös oka. Megadható már pusztán két korreláló párt tartalmazó egyszerű mértékter is, amelynek nem létezik *közös* közös okot tartalmazó kiterjesztése. A *közös* közös okkal való kiterjeszthetőség szükséges és elégséges feltételei viszont nem ismertek. A *közös* közös ok fogalma tehát radikálisan erősebb fogalom, mint a közös ok fogalma. Ennélfogva különböző korrelációk nem feltétlenül magyarázhatók egyetlen *közös* közös okkal, bármennyire finomítjuk is a világról alkotott leírásunkat.

A reichenbachi közös ok elvvel szemben azonban a fenti tétel nem jelent ellenvetést, mivel az elv – legalábbis eredeti formájában – nem vonatkozik több korrelációhoz tartozó *közös* közös okokra, pusztán egyetlen korrelációhoz tartozó közös okra vonatkozik.

## A KÖZÖSOK-RENDSZER

A reichenbachi közös ok definícióval szemben több ellenvetés is megfogalmazható (E. Szabó 2002, 5.5 fejezet). Ezek egyike szerint a reichenbachi definíció szükségtelenül restriktív, és csak pozitív korrelációkra épül. Egyrészt egy  $A$  és  $B$  esemény közötti *negatív* korreláció ugyanúgy kauzális magyarázatot követel, mint egy pozitív korreláció (vö. Gyenis B. – Rédei 2004). Másrészt az  $A$  és  $B$  esemény közötti – akár negatív, akár pozitív – korreláció létrejöttében az  $A$  és  $C$  esemény közötti, illetve  $B$  és  $C$  esemény közötti pozitív és negatív korrelációk ugyanúgy szerepet játszhatnak. Megmutatható (E. Szabó 2002, 125), hogy a (3)–(4) feltételek jelentős gyengítése mellett is ugyanúgy triviálisan következik az  $A$  és  $B$  közötti korreláció, azzal a különbséggel, hogy a korreláció lehet negatív is. Vagyis, a deduktív-nomologikus magyarázati modell értelmében, a közös ok jelenléte ebben az esetben is *magyarázza* az  $A$  és  $B$  közötti korrelációt.

A másik ellenvetés, hogy a reichenbachi értelemben definiált közös ok alkalmatlannak tűnik azokban az esetekben, amikor egy korreláció létrejöttében több kauzális faktor játszik egyszerre szerepet. Ha például elgondolunk két olyan  $C$  és  $C'$  eseményt, melyek bekövetkezése külön-külön is megnöveli az  $A$  és  $B$  események együttes bekövetkezési valószínűségét, akkor teljesen értelmetlen megkövetelnünk, hogy akár  $C$ , akár  $C'$  teljesítse a negáltra vonatkozó

(2) árnyékolási tulajdonságot. Hiszen ha  $C$  nem következik be, a korrelációnak még nem kell eltűnnie, ugyanis  $C'$  ettől még bekövetkezhet.

Végül meg kell említenünk, hogy számos egyszerű példa mutatható, amelyben az eseményalgebra sok (akár kontinuum sok) a reichenbachi definíciót kielégítő eleme van, mégis teljesen kontraintuitív volna ezek bármelyikét közös oknak tekintenünk, míg az adott példában nyilvánvaló közös ok nem teljesíti a reichenbachi kondíciókat.

Mindezek tükrében úgy tűnik, hogy a Reichenbach által definiált közösok-fogalom nem tükrözi helyesen azt a fogalmat, amellyel egy korrelációt intuitíve elfogadhatóan magyarázni tudunk. Nem tekinthető véletlennek tehát, hogy a kauzálisan szeparált események közötti korrelációk közös ok típusú magyarázatában, jelesül a már említett kvantummechanikai kísérletekben tapasztalt korrelációk rejtett paraméteres magyarázatában használt, a fizikus intuíciónak jobban megfelelő közösok-fogalom hasonlít, de *nem egyezik meg* a Reichenbach által definiált közös ok fogalmával. A következőkben a közös oknak ezt az intuitív fogalmát fogjuk pontosan, az eddig megismert kontextusba helyezve értelmezni.

Egy  $(A, B)$  korreláló pár *közösok-rendszere* alatt az  $(X, \Sigma, p)$  mértéktérnek egy olyan  $\{C_i\}_{i=1, \dots, n}$  partícióját értjük, amelynek minden  $C_i$  tagjára teljesül a következő összefüggés:

$$p(A \cap B | C_i) = p(A | C_i) \cdot p(B | C_i). \quad (5)$$

A fizikában használt rejtett paraméter fogalma pontosan a fent definiált fogalomnak felel meg: a közösokrendszer egy  $C_i$  tagja az az eseménytípus, amelyet az jellemez, hogy a rejtett paraméter egy meghatározott értéket vesz fel.

A közösok-rendszer fogalma mögött a kauzalitás természetére vonatkozóan azon intuitív kép húzódik meg, hogy a szóban forgó korrelációt nem feltétlenül egyetlen esemény hozza létre, hanem esetleg események olyan rendszere, amelyek egymással összekapcsolódva fejtik ki kauzális hatásukat az okozatokra nézve. A rendszer elemei tehát a reichenbachi négy kondícióból csak (1)-et teljesítik. Vagyis a közösok-rendszer fogalma pontosan orvosolja a Reichenbach-féle közösok-fogalommal szemben felmerült kifogásokat. Speciálisan, egy kételemű közösok-rendszer éppen megegyezik a reichenbachi közös ok fogalmával, elhagyva belőle a (3)–(4) egyenlőtlenségeket.

Mi a helyzet tehát a reichenbachi közös ok elvére vonatkozó fenti tételekkel, ha a reichenbachi közös ok fogalmát a közösok-rendszer fogalmával cseréljük fel? Ebben az esetben nemcsak a közösok-kiterjeszthetőségre vonatkozó tételek maradnak továbbra is érvényben, hanem a *közös* közösok-kiterjeszthetőségre vonatkozó tételek is érvényesek lesznek az alábbi értelemben: tetszőleges  $(X, \Sigma, p)$  valószínűségi mértéktérben korrelációk egy tetszőleges véges halmazhoz található olyan  $n$  természetes szám, hogy  $(X, \Sigma, p)$  kiterjeszthető úgy, hogy a kiterjesztett mértéktérben a korrelációs halmaz minden elemének van egy  $n$  elemű *közös* közösok-rendszere.

A teljesség kedvéért érdemes megemlítenünk, hogy ha a közösok-rendszer fogalmát olyan módon módosítjuk, hogy (5) mellett megköveteljük a (3)–(4) egyenlőtlenségek általánosításának tekinthető

$$[\rho(A|C_i) - \rho(A|C_j)] \times [\rho(B|C_i) - \rho(B|C_j)] > 0 \quad (i \neq j)$$

feltételt, amely garantálja az  $A$  és  $B$  közötti korreláció pozitivitását, akkor a kiterjeszhetőségre vonatkozó pozitív, illetve a *közösok*-kiterjeszhetőségre vonatkozó negatív tételek továbbra is érvényben maradnak (Hofer-Szabó – Rédei 2004, 2006). Az ilyen extra tulajdonságú közösok-rendszereket hívjuk *reichenbachi közösok-rendszernek*.<sup>5</sup>

## NEM KLASSZIKUS MÉRTÉKTEREK

Az uralkodó nézet szerint a kvantumjelenségek értelmezéséhez egy nem klasszikus valószínűségelmélet bevezetése szükséges, ahol is az eddigi vizsgálódások alapjául szolgáló  $(X, \Sigma, p)$  mértéktér helyébe egy kvantumvalószínűségi mértéktér lép.<sup>6</sup> A kvantumvalószínűségi mértéktérben természetesen a közös ok definíciójában szereplő fogalmakat is adaptálni kell az új formalizmushoz, de ez minden további nélkül megtehető. Ezek után bizonyíthatóvá válik a klasszikus kiterjeszhetőségi tétel kvantummos megfelelője: *Tetszőleges* kvantumvalószínűségi mértéktérre igaz, hogy a mértéktér korrelációinak *tetszőleges* megszámlálhatóan végtelen halmazához létezik a mértéktérnek olyan kiterjesztése, amelyben a halmazba tartozó korrelációk mindegyikének van közös oka (Hofer-Szabó – Rédei – E. Szabó 1999). Mint látjuk, a reichenbachi közös ok elv teljesülésének nincs valószínűség-elméleti akadály a kvantumvalószínűség-elmélet keretein belül sem. Amennyiben a reichenbachi közös ok elvet cáfolni akarjuk, a közös ok definíciójába a felsorolt négy kritériumon túl extra feltevéseket is be kell illesztenünk.

Honnan jöhetnek ilyen járulékos feltevések? Nyilvánvalóan a fizikai szituáció további jellemzéséből. A legfontosabb ilyen jellemző az események tér-időbeli helyzete. Ha ezt a téridőbeli elhelyezkedést sikerül modelleznünk a

<sup>5</sup> A részletek mellőzésével megjegyezzük, hogy nemcsak a korreláló események közös okát lehet eseményről partícióra „finomítani”, hanem magukat a korreláló eseményeket is, sőt, a korreláció korábban bevezetett definícióját is általánosíthatjuk, ahogyan ez a statisztikában a különféle asszociációs mutatószámok bevezetésekor történik. A reichenbachi közös ok és a közösok-rendszer így speciális eseteivé válnak egy úgynevezett *általánosított reichenbachi közös oknak*; a kiterjeszhetőséggel és zártsággal kapcsolatos eredmények egy részét is lehet ezen általános keretek között igazolni (Gyenis B. – Rédei 2010).

<sup>6</sup> Nem feltétlenül kell egyetértenünk ezzel a nézettel. A szerzők véleménye is megoszlik ebben a kérdésben (vö. Rédei 1998; E. Szabó 2001, 2002). A kvantumvalószínűség-elmélet alapjairól lásd Rédei 1998 vagy E. Szabó 2002.

valószínűségelmélet keretein belül, akkor további megszorításokat nyerhetünk a korrelációk magyarázataként szóba jöhető közös okok körére nézve, és ez új fénybe helyezheti a közös ok elv érvényességére vonatkozó eddigi megfontolásainkat is.

A kvantumelméleti események téridőbeli *locus*át is figyelembe vevő elmélet a kvantumtérelmélet, amely a megfigyelhető mennyiségeket téridőbeli tartományokhoz rendeli.<sup>7</sup> Ez az elmélet korrelációt jósol térszerűen szeparált, azaz a relativitáselmélet szerint kauzálisan független tartományokhoz tartozó események között. Ezek a korrelációk tehát nem magyarázhatók a korreláló események közötti közvetlen kauzális hatás eredményeként. Ha a reichenbachi közös ok elvet – ebben a lokalitással kiegészített formában – érvényben kívánjuk tartani, nem marad más lehetőségünk, minthogy a korrelációt egy, a téridőben megfelelően elhelyezkedő közös ok segítségével magyarázzuk. A „megfelelően elhelyezkedő” itt a relativitáselmélet szellemében azt jelenti, hogy a közös ok téridőbeli helye a korreláló események múltbeli fénykúpjainak metszetében van, vagyis a téridőnek abban a tartományában, ahonnan legfeljebb fénysebességgel terjedő kauzális hatások érkehetnek a két eseményhez. A közös oknak ez a lokalizációja azonban nem az egyetlen lehetőség; mind gyengébb, mind erősebb lokalizáció elképzelhető. A gyengébb lokalizáció megelégszik egy olyan tartománnyal, amely a két korreláló esemény közül legalább az egyiket kauzálisan befolyásolni tudja, vagyis a korreláló események múltbeli fénykúpjainak uniójával. Az erősebb lokalizáció ezzel szemben egy olyan tartományt jelent, amelynek minden téridőbeli pontja mindkét korreláló esemény minden pontját kauzálisan befolyásolni képes. Ennélfogva a reichenbachi közös ok elv három nem ekvivalens formában implementálható a kvantumtérelméletbe. Így azután ismét csak felvethető a kérdés, hogy a kvantumtérelmélet kauzálisan elég gazdag-e ahhoz, hogy a térszerűen szeparált korrelációkhoz a három különböző értelemben reichenbachi közös okot szolgáltasson. Kiderült, hogy a közös ok elv gyenge értelemben fenntartható, azaz térszerűen szeparált korreláló eseményekhez mindig található közös ok az események múltbeli fénykúpjainak uniójában (Rédei – Summers 2005). Kiderült továbbá az is, hogy az erős közös ok elv megsérthető az ún. *wedge* tartományokon megadható korrelációk segítségével. A harmadik, köztes értelemben vett (és egyben legfontosabb) reichenbachi közös ok elv érvényességéről jelenleg azonban keveset tudunk.

Hangsúlyozzuk, hogy a fentiekben a reichenbachi közös ok elvet az eredeti értelemben vettük, ahol a korrelációkat közös *okokkal*, nem pedig közösok-*rendszerekkel* kívánjuk magyarázni. Hogy a nem klasszikus esetben mi a helyzet a közösok-rendszerrel kapcsolatban, az egyelőre nyitott kérdés.

<sup>7</sup> A kvantumtérelméletben az események mint speciális fizikai mennyiségek a lokális Neumann-algebrák projektoraival vannak reprezentálva.

## AZ EPR–BELL-PARADOXON

A reichenbachi közös ok elvvel szembeni legfőbb kihívást az ún. EPR–Bell-paradoxon jelenti.<sup>8</sup> A fizikai szituáció,<sup>9</sup> amelyet a nyolevanas évektől kezdve kísérletileg is ellenőriztek, a következő.

Egy részecskeforrásból megfelelően preparált állapotú kvantumrészecskék repülnek szét jobbra és balra. A szétrepülő részecskéken ezután, egymástól távol, egy-egy spinvetületmérést hajtunk végre, melyeknek a kvantummechanika törvényei szerint két lehetséges kimenetele van: egy adott irányban mért spinvetület lehet *pozitív* vagy *negatív*. A mérés mindkét részecske esetében két-két tetszőleges irányban történhet, amely irányokat (legalábbis a legújabb mérésekben) a részecskék repülési ideje alatt egymástól független random kapcsolók választják meg. Egy kísérleti futamban tehát mindkét oldalon egy-egy mérési irányt és egy hozzá tartozó pozitív vagy negatív spinértéket regisztrálnak. Jelöljük a bal és jobb oldali két-két mérési irányt rendre  $a_i$ -val és  $b_j$ -vel ( $i, j = 1, 2$ ), a megfelelő irányú spinmérés *pozitív* kimeneteleit pedig rendre  $A_i$ -val és  $B_j$ -vel. Az ismételt mérések során megállapíthatjuk a különböző mérési irányok választásának, valamint a mérések kimeneteleinek statisztikáját. Az egyszerűség kedvéért egy tipikus, ilyen mérési elrendezésben mért konkrét eredményeket adunk meg példaként:

$$p(a_i) = p(b_j) = \frac{1}{2} \quad (6)$$

$$p(A_i) = p(B_j) = \frac{1}{4} \quad (7)$$

$$p(a_i \cap b_j) = p(a_i) p(b_j) \quad (8)$$

$$p(a_i \cap B_j) = p(a_i) p(B_j) \quad (9)$$

$$p(A_i \cap b_j) = p(A_i) p(b_j) \quad (10)$$

$$p(A_1 \cap B_1) = p(A_1 \cap B_2) = p(A_2 \cap B_2) = \frac{3}{32} \quad (11)$$

$$p(A_2 \cap B_1) = 0 \quad (12)$$

<sup>8</sup> Einstein–Podolsky–Rosen 1935; Bell 1964. (Az EPR paradoxon részletesebb áttekintését lásd E. Szabó 2002, 8.4–8.6 fejezet; E. Szabó 2008.)

<sup>9</sup> Bohm–Aharonov 1957.

Mint láthatjuk, a jobb és bal oldali mérések kimenetelei között korreláció van:

$$p(A_1 \cap B_1) - p(A_1)p(B_1) = \frac{1}{32}, \quad (13)$$

$$p(A_1 \cap B_2) - p(A_1)p(B_2) = \frac{1}{32}, \quad (14)$$

$$p(A_2 \cap B_2) - p(A_2)p(B_2) = \frac{1}{32}, \quad (15)$$

$$p(A_2 \cap B_1) - p(A_2)p(B_1) = \frac{-1}{16}. \quad (16)$$

A kérdés ezek után az, hogy mi a kauzális magyarázata a szóban forgó négy korrelációnak.

A modern kísérleti technika segítségével lehetővé vált, hogy a két mérést olyan távolságban végezzék el, illetve a mérési időablakokat olyan rövidnek válasszák, hogy ezáltal garantálható legyen a két mérés kauzális szeparációja. Pontosabban a két mérőhelyen történtek így csak akkor lehetnének egymásra kauzális hatással, ha ez a hatás a fénysebességnél nagyobb sebességgel terjedne; az ilyen szuperlumináris hatásokat azonban más fizikai megfontolások alapján ki szoktuk zárni. Ha azonban a közvetlen kauzális hatást kizártuk, akkor a reichenbach-i közös ok elve már csak úgy teljesülhet, ha a korrelációk valamilyen közös okból származnak.

Mint említettük, a fizikai irodalomban az EPR-kísérletben tapasztalt korrelációk közös ok típusú magyarázata alatt egy rejtett paraméteres magyarázatot szokás érteni, ami az általunk bevezetett fogalmak szerint egy *közös* közösok-rendszerrel történő magyarázatnak felel meg. A kérdés tehát az, hogy beágyazhatók-e az EPR-kísérletben megfigyelt események a kísérletben megfigyelt (6)–(12) relatív gyakoriságokkal egy olyan klasszikus valószínűségi elméletbe, amelyben a spinmérések eredményei között fennálló (13)–(16) korrelációknak *közös* közösok-rendszere van? A kiterjesztési tételek alapján a válasz az lenne, hogy igen. Felmerül azonban egy további követelmény a közös okkal szemben, amelyről eddig még nem tettünk említést.

A ténylegesen elvégzett mérésekben a mérési irányok választása a jobb és a bal oldalon két egymástól független random kapcsolóberendezéssel történik. Az egymástól való függetlenség *nem hipotézis, hanem megfigyelt tény*, melyet a (8)-as egyenlet fejez ki. Mármost plauzibilisnek tekinthető azt is feltételeznünk, hogy a random kapcsolók működése minden mástól is független, jelesül független a  $\{C_k\}_{k=1 \dots n}$  közösok-rendszerbe tartozó eseményektől. Azaz feltesszük, hogy

$$\begin{aligned} p(a_i \cap C_k) &= p(a_i)p(C_k) \\ p(b_j \cap C_k) &= p(b_j)p(C_k) \end{aligned} \quad (i, j = 1, 2; k = 1 \dots n) \quad (17)$$



Ez az úgynevezett *konspirációmentesség* tehát egy metafizikai *feltevés*, melyet intuitív argumentumokkal szokás alátámasztani. Hármat említünk meg:

1. A közösok-rendszer  $C_k$  eseményei feltételezhetően olyan fizikai események, amelyek a jobb és bal oldali mérések hátrafénykúpjainak metszetében történnek, hiszen a  $C_k$  eseményeknek kauzális hatással kell lenniük a mérés kimeneteleire, és feltevésünk szerint nem létezik szuperluminális kauzális hatás. Mivel pedig a random kapcsolók működését a jobb és a bal oldalon lokális random események irányítják, melyek kívül esnek a két hátrafénykúp metszetén, így fizikailag nem plauzibilis feltételezni egy olyan rejtett konspirációt, amely szerint a részecskék spinjeit meghatározó közös okok valamilyen újabb közös ok típusú magyarázatra szoruló korrelációban állnának a kapcsolókban történő random eseményekkel.
2. Elvben a random kapcsolásokat vezérelhetnénk az univerzum két ellenkező végéből jövő, nagyon távoli random eseményekkel (pulzárak stb.) is, és így szintén nem volna plauzibilis feltenni, hogy a világot olyan rejtett konspiráció hatja át, amely összehangolja a sok milliárd fényévre lévő különböző random történéseket a laboratóriumunkban keltett részecskék spinméréseiben jelentkező eseményeket meghatározó közös okokkal.
3. Elvben a random kapcsolók helyén ülhetne két laboráns is, akik szabad akaratukból döntenek, hogy melyik mérést válasszák. Nem volna tehát plauzibilis feltételeznünk, hogy a laboránsok szabad választásaira a részecskék viselkedését meghatározó közös okok hatással lennének.

Jó okunk van tehát feltenni, hogy a közös közösok-rendszer kielégíti a (17) feltételt. E feltevés mellett azonban a *közös* közösok-rendszer létezéséből bizonyos egyenlőtlenségek vezethetők le, az úgynevezett Bell-egyenlőtlenségek, melyeket a kísérletben tapasztalt relatív gyakoriságoknak kellene teljesíteniük. A mért relatív gyakoriságok azonban sértik a Bell-egyenlőtlenségeket (lásd például E. Szabó 2000). A konklúzió tehát az, hogy a konspirációmentesség feltételét kielégítő *közös* közösok-rendszerrel az EPR-kísérletben tapasztalt korrelációk nem magyarázhatók.

Az egyik lehetséges kiút, hogy a korrelációk magyarázatára nem *közös* közösok-rendszer létezését, hanem úgynevezett *külön* közösok-rendszerek létezését tesszük fel. Megmutatható (E. Szabó 2000), hogy az EPR-kísérletben megfigyelt események a kísérletben megfigyelt (6)–(12) relatív gyakoriságokkal beágyazhatók egy olyan klasszikus valószínűségi elméletbe, amelyben a spinmérések eredményei között fennálló (13)–(16) korrelációknak külön-külön létezik közösok-rendszere úgy, hogy a közös okok kielégítik a (17) feltételt. Ez az eredmény azt sugallja, hogy az EPR-kísérlet nem cáfolja a Reichenbach-féle közös ok elvet. A helyzet azonban bonyolultabb. A közös okok ugyan teljesítik a konspirációmentességet megfogalmazó (17) feltételt, a modellben azonban újabb konspirációt jelentő korrelációk lépnek fel: a mérésválasztások korrelálnak a

közös okokból az eseményalgebra műveleteivel képzett összetett eseményekkel. Komputeres vizsgálatok arra a sejtésre engedtek következtetni, hogy nem létezik az EPR-korrelációknak konspirációmentes *külön* közösok-rendszereket megengedő modellje (E. Szabó 2000). Bizonyos speciális kondíciók mellett a sejtésre vannak bizonyítások (Grasshoff–Portmann–Wüthrich 2005; Portmann–Wüthrich 2007; Hofer-Szabó 2008, 2010).

Az EPR-kísérlet – és néhány hasonló spinkorrelációs kísérlet – tehát valóban kihívást jelent a Reichenbach-féle közös ok elv számára. Valójában ez az egyetlen olyan szituáció, amikor az elv sérülni látszik. Érdemes megjegyeznünk, hogy annak ellenére, hogy a kvantummechanika kontextusában merül fel, az EPR–Bell-probléma teljesen független a kvantummechanikától. Laboratóriumi kísérletekben megfigyelt makroszkopikus események relatív gyakoriságairól van szó, melyekre nincs kauzális magyarázat. Ebből a szempontból teljesen mellékes az a körülmény, hogy a kísérletben megfigyelt relatív gyakoriságok megegyeznek a kvantummechanika jóslataival. A probléma feloldása tehát csak két úton képzelhető el: vagy a laboratóriumi mérések közvetlen értelmezésére vonatkozó új megfontolásokkal (Fine 1982, 1991; Larsson 1999a, 1999b; E. Szabó 2000b; E. Szabó – Fine 2002), vagy a kauzális fogalmainkat újraértelmező, illetve pontosító, részben metafizikai megfontolásokkal (E. Szabó 1989; Rédei 1995; Belnap – E. Szabó 1996; Rédei–Summers 2010; Rédei 2010).

## ÖSSZEFOGLALÁS

A reichenbachi közös ok elvre, vagyis korrelációk kauzális magyarázatára vonatkozóan tehát a következő tézisszerű állításokat tehetjük:

1. A reichenbachi közös ok elv érvényessége nem dönthető el fizikai szituációk intuitív érvelésre támaszkodó, informális megközelítésén keresztül; az elv elemzése a közös ok fogalmának pontos valószínűségi modellezését előfeltételezi.
2. A formális megközelítés egyik legfontosabb eredménye, hogy minden fizikai szituációt reprezentáló klasszikus és kvantumos valószínűségi mértékter kiterjeszhető úgy, hogy bármely benne szereplő korrelációnak legyen közös oka.
3. Mi több, ez a kiterjesztés úgy is elvégezhető, hogy a bővebb mértékterben egyáltalán ne legyen közös ok nélküli korreláció. A közös ok típusú magyarázatra szoruló korrelációk köre továbbá tetszés szerint hangolható egy  $R_{ind}$  kauzális függetlenségi reláció segítségével.
4. A közös ok elvben szereplő közös okot szigorúan meg kell különböztetni a *közös* közös ok fogalmától, több korreláció egyazon közös okától. Ez utóbbira vonatkozóan a valószínűségi mértékterek általában nem terjeszthetők ki.

5. Amennyiben közös okról közösok-rendszerre térünk át, úgy nemcsak a közösok-kiterjeszhetőségre vonatkozó tételek maradnak érvényben, hanem a *közös* közösok-kiterjeszhetőségre vonatkozó tételek is érvényesek lesznek a fent vázolt értelemben.
6. A klasszikusról kvantumvalószínűségi mértéktérre térve át a közösok-kiterjeszhetőségre vonatkozó tétel szintén érvényben marad.
7. A kvantumtérelméletben a közös ok múltbeli lokalizációjától függően legalább háromféleképpen megfogalmazhatjuk a közös ok elvét. Ezek közül egy biztosan igaz, egy biztosan hamis, egyről (a legfontosabbról) pedig keveset tudunk.
8. Végül a reichenbachi közös ok elv az eredetitől legtávolabb eső megfogalmazását az EPR-korrelációk kauzális magyarázatánál nyerte el, mivel itt az elv jelentését megterheltek egyfelől azok az extra valószínűségi megszorítások, amelyek a korreláló események, a mérésválasztások, illetve a közös ok téridőbeli elhelyezkedéséből adódtak, másfelől a keresés mindvégig *közös* közösok-rendszerre vonatkozott, nem pedig egyszerűen közös okra. Az ily módon megterhelt közös ok elv érvényességét végül a Bell-egyenlőtlenségek kizárják. Hogy mi a helyzet akkor, ha a *közös* közösok-rendszert *külön* közösok-rendszerrel helyettesítjük, egyelőre nem ismeretes.

## IRODALOM

- Belnap, Nuel – László E. Szabó 1996. Branching Space Time Analysis of the GHZ Theorem. *Foundations of Physics*. 26. 989–1002.
- Bohm, David – Yakir Aharonov 1957. Discussion of Experimental Proof for the Paradox of Einstein, Rosen, and Podolsky. *Physical Review*. 108. 1070–1076.
- Einstein, A. – B. Podolsky – N. Rosen 1935. Can Quantum-Mechanical Description of Physical Reality be Considered Complete? *Physical Review*. 47. 777–780.  
(A cikk elérhető az interneten: [http://prola.aps.org/pdf/PR/v47/i10/p777\\_1](http://prola.aps.org/pdf/PR/v47/i10/p777_1))
- E. Szabó, László 1989. Quantum Causal Structure and the Einstein–Podolsky–Rosen Experiment. *International Journal of Theoretical Physics*. 28. 35–47.
- E. Szabó, László 2000a. On an Attempt to Resolve the EPR–Bell Paradox via Reichenbachian Concept of Common Cause. *International Journal of Theoretical Physics*. 39. 901–911.
- E. Szabó, László 2000b. On Fine’s Resolution of the EPR–Bell Problem. *Foundations of Physics*. 30. 1891–1909.
- E. Szabó, László 2001. Critical Reflections on Quantum Probability Theory. In Rédei Miklós – Michael Stoeltzner (szerk.) *John von Neumann and the Foundations of Quantum Physics*. Kluwer, Dordrecht.
- E. Szabó, László – Arthur Fine 2002. A Local Hidden Variable Theory for the GHZ Experiment. *Physics Letters*. A295. 229–240.
- E. Szabó László 2002. *A nyitott jövő problémája – véletlen, kauzalitás és determinizmus a fizikában*. Budapest, Typotex.
- E. Szabó, László 2008. The Einstein–Podolsky–Rosen Argument and the Bell Inequalities. *Internet Encyclopedia of Philosophy*.  
URL: <http://www.iep.utm.edu/epr> Hozzáférés: 2010. 10. 01.

- Fine, Arthur 1982. Some Local Models for Correlation Experiments. *Synthese*. 50. 279–294.
- Fine, Arthur 1991. Inequalities for Nonideal Correlation Experiments. *Foundations of Physics*. 21. 365–378.
- Gyenis, Balázs – Miklós Rédei 2004. When Can Statistical Theories Be Causally Closed? *Foundations of Physics*. 34. 1285–1303.
- Gyenis, Balázs – Miklós Rédei 2010. Causal Completeness of Generalized Probability Theories. In Mauricio Suárez (szerk.) *Probabilities, Causes and Propensities in Physics*. Dordrecht, Springer.
- Gyenis, Zalán – Miklós Rédei 2010. Characterizing Common Cause Closed Probability Spaces. URL.: <http://philsci-archive.pitt.edu/archive/00005390/> Hozzáférés: 2010. 10. 31.
- Grasshoff, Gerd – Samuel Portmann – Adrian Wüthrich 2005. Minimal Assumption Derivation of a Bell-type Inequality. *The British Journal for the Philosophy of Science*. 56. 663–680.
- Hofer-Szabó, Gábor – Miklós Rédei – László E. Szabó 1999. On Reichenbach's Common Cause Principle and on Reichenbach's Notion of Common Cause. *The British Journal for the Philosophy of Science*. 50. 377–399.
- Hofer-Szabó, Gábor – Miklós Rédei – László E. Szabó 2002. Common Causes are not Common Common Causes. *Philosophy of Science*. 69. 623–633.
- Hofer-Szabó, Gábor – Miklós Rédei 2004. Reichenbachian Common Cause Systems. *International Journal of Theoretical Physics*. 43. 1819–1826.
- Hofer-Szabó, Gábor – Miklós Rédei 2006. Reichenbachian Common Cause Systems of Arbitrary Finite Size Exist. *Foundations of Physics*. 35. 745–756.
- Hofer-Szabó, Gábor 2008. Separate Versus Common-Common-Type Derivations of the Bell Inequalities. *Synthese*. 163/2. 199–215.
- Hofer-Szabó, Gábor 2010. Bell( $\delta$ ) Inequalities Derived from Separate Common Causal Explanation of Almost Perfect Anticorrelations. *Foundations of Physics* (megjelenés alatt).
- Larsson, Jan-Åke 1999a. Modeling the Singlet State with Local Variables. *Physics Letters*. A 256. 245–252.
- Larsson, Jan-Åke 1999b. Detector Efficiency in the Greenberger–Horne–Zeilinger Paradox: Independent Errors. *Physical Review*. A 59. 4801–4804.
- Pearl, Judea 2000. *Causality: Models, Reasoning, and Inference*. Cambridge, Cambridge University Press.
- Portmann Samule – Adrian Wüthrich 2007. Minimal Assumption Derivation of a Weak Clauser–Horne Inequality. *Studies in History and Philosophy of Modern Physics*. 38. 844–862.
- Price, Huw 1996. *Time's Arrow & Archimedes' Point: New Directions for the Physics of Time*. New York, Oxford University Press.
- Reichenbach, Hans 1956. *The Direction of Time*. Berkeley, University of California Press.
- Rédei, Miklós 1995. Logical Independence in Quantum Logic. *Foundations of Physics*. 25. 411–422.
- Rédei, Miklós 1998. *Quantum Logic in Algebraic Approach* (Fundamental Theories of Physics Vol. 91.) Dordrecht–Boston–London, Kluwer Academic Publishers.
- Rédei, Miklós 2010. Operational Separability and Operational Independence in Algebraic Quantum Mechanics. *Foundations of Physics*. 40. 1439–1449.
- Rédei, Miklós – Stephen J. Summers 2002. Local Primitive Causality and the Common Cause Principle in Quantum Field Theory. *Foundations of Physics*. 32. 335–355.
- Rédei, Miklós – Stephen J. Summers 2007. Remarks on Causality in Relativistic Quantum Field Theory. *International Journal of Theoretical Physics*. 46. 2053–2062.
- Rédei, Miklós – Stephen J. Summers 2010. When are Quantum Systems Operationally Independent? *International Journal of Theoretical Physics*. 49. 3250–3261.
- Salmon, Wesley C. 1978. Why ask 'why'? *Proceedings and Addresses of the American Philosophical Association*. 51/6. 683–705.

- Sober, Eliot 1988. The Principle of the Common Cause. In James H. Fetzer (szerk.) *Probability and Causality*. Dordrecht, Reidel. 211–228.
- Spirtes, Peter – Clark Glymour – Richard Scheines 2000. *Causation, Prediction, and Search* (2. kiadás). Cambridge/MA, MIT Press.
- Szabó Gábor 2006. A reichenbachi közös ok metafizikája. *Világosság*. 47/5. 87–94.
- Van Fraassen, Bas C. 1982. Rational Belief and Common Cause Principle. In Robert McLaughlin (szerk.). *What? Where? When? Why?* Dordrecht, Reidel. 193–209.
- Van Fraassen, Bas C. 1989. The Charybdis of Realism: Epistemological Implications of Bell's Inequality. In James T. Cushing – Ernan McMullin (szerk.) *Philosophical Consequences of Quantum Theory*. Indiana, University of Notre Dame Press. 97–113.

## Távolba ható okok

### I. BEVEZETÉS

Akármit gondolunk is az okság metafizikai természetéről általában, tagadhatatlannak tűnik, hogy oksági kijelentéseink jó része valamilyen relációra, viszonyra utal. Ezért az okság metafizikájának alapvető kérdése egyrészt az, hogy milyen típusú létezők állhatnak egymással oksági kapcsolatban, másrészt pedig az, hogy mi teszi a köztük fennálló kapcsolatot oksági kapcsolattá. E két kérdésre adott válasz természetesen nem független egymástól. Az oksági viszonyban álló entitások jellemzése nyilvánvalóan befolyásolja azt, hogy miként jellemezzük a köztük fennálló relációkat.

Érdekes módon azonban az oksági viszony formális tulajdonságait tárgyaló kortárs irodalom egy jelentős része az egymással oksági kapcsolatban álló entitások releváns tulajdonságainak vizsgálata nélkül kísérli meg értelmezni az oksági viszony természetét. Mint látni fogjuk, ennek magyarázata abban a hume-i hagyományban keresendő, amely szerint oksági viszony csak „elkülönült létezők” (*distinct existences*) között állhat fenn. Filozófiai és filozófiatörténeti viták tárgya, mit is értett valójában Hume „elkülönültségen”, kézenfekvőnek látszik azonban az az értelmezés, mely szerint oksági kapcsolatban csak olyan létezők állhatnak egymással, amelyek sem logikai, sem pedig metafizikai értelemben nem határozzák meg egymás azonosításának feltételeit. Ebből pedig az következik, hogy az oksági kapcsolatban álló entitások közti viszony természetének tisztázása független kell hogy legyen attól, miként azonosítjuk ezeket az entitásokat. Másképp fogalmazva: azt a kérdést, hogy az oksági viszony milyen formális tulajdonságokkal rendelkezik, függetlenül kell kezelnünk attól a kérdéstől, hogy mit és miért tekintünk okságilag releváns tulajdonságnak.

A következőkben egy alternatív megközelítést javaslok az oksági viszonyok formális jellegzetességeinek megértésére.<sup>1</sup> Mindenekelőtt amellet fogok érvelni, hogy az oksági viszony formális tulajdonságaira vonatkozó kérdés nem választható el attól a kérdéstől, hogy mit tekintünk okságilag releváns tulajdon-

<sup>1</sup> E jellegzetességekről, illetve a velük kapcsolatos problémákról igen jó rövid összefoglalót ad Molnar 2003. 190–194.

ságnak. Bár úgy vélem, hogy javaslatom kiterjeszthető más jellegzetességekre is, részletesen csak a tranzitivitás kérdésével foglalkozom. Ennek oka, hogy feltevésem szerint a tranzitivitás „formális” problémájának vizsgálata hozzásegíthet bennünket az oksággal kapcsolatos számos más metafizikai kérdés jobb megértéséhez is. Ezek közül talán a legérdekesebb a kauzálisan releváns tulajdonságok egymás közti viszonyára vonatkozik, vagyis arra, hogy ezek a tulajdonságok redukálhatók-e valamilyen módon egymásra; továbbá arra, hogy ha válaszunk nemleges, mi magyarázza a redukció lehetetlenségét.

A tanulmányban először röviden szót ejtek az oksági reláció reflexivitásával és aszimmetriájával kapcsolatos kérdéstről. Ez utóbbi két szempontból is releváns. Egyrészt, mivel az okság aszimmetriájának hume-iánus magyarázatai, úgy tűnik, megkövetelik a tranzitivitást. Másrészt, mivel egy fontos tekintetben – az oksági viszony perspektivikusságának kérdésében – érveim rokonságban állnak egy az okság aszimmetriáját magyarázó nagy hatású és a hagyományos hume-i megközelítésekkel szemben álló elmélettel. A következő részben azt mutatom meg, miért fontos az okság tranzitivitásának feltételezése a hume-iánus okságelméletek, mindenekeelőtt a David Lewis által megfogalmazott kontrafaktuális elmélet számára. Ezután röviden ismertetem a kontrafaktuálisok és a tranzitivitás kapcsán kialakult kurrens vitát. Végül, utóbbiból kiindulva, az okság tranzitivitásával kapcsolatos saját elképzelésem mellett érvelek, és igyekszem megmutatni, milyen következtetések adódnak ebből az oksági redukció problémájára nézve.

## II. IRREFLEXIVITÁS ÉS ASZIMMETRIA

Annak ellenére, hogy – mint említettem, és mint arra a későbbiek során még részletesen ki fogok térni – a kortárs, jórészt hume-i indíttatású megközelítés rendkívül nehézé teszi az oksági viszonyok formális tulajdonságainak tisztázását, bizonyos kérdésekben egyetértés uralkodik. A konszenzus, nem véletlenül, éppen a tranzitivitás kérdése kapcsán szűnik meg.

Szinte minden általam ismert szerző egyetért abban, hogy az oksági reláció irreflexív, tehát hogy „semmi sem lehet önmaga oka”. Ez rendkívül meggyőzően hangzik, ha az okságot a létezés terminusaiban értelmezzük, tehát ha az irreflexivitást azt fejezi ki, hogy sem egy tárgy, sem egy esemény nem lehet önmaga létezésének oka. Van azonban a reflexív okozás lehetőségével kapcsolatban egy olyan kérdés, amely releváns e tanulmány tárgya szempontjából is. Talán nem elképzelhetetlen, hogy léteznek olyan „okszági hurkok”, melyek során egy esemény, egy vagy több tőle különböző esemény közvetítésével, önmagát okozza (Lewis 1983c. 212–213). Egyesek szerint a feltételezés bizarr, sőt logikailag lehetetlen (Mellor 1995). Ha azonban lehetséges ilyen okozás, az csakis egy tranzitív oksági láncon keresztül valósulhat meg. A reflexív okozás lehetősége tehát a tranzitivitás problémájához vezet bennünket.

Ugyancsak nagy egyetértés uralkodik az oksági reláció aszimmetrikus voltának tekintetében: ha  $C$  oka  $A$ -nak, lehetetlen, hogy  $A$  is oka legyen  $C$ -nek. Azonban talán ez sem magától értetődő. Végül is az oksági viszony egyik elemi formája az *interakció*, ami pedig éppen azt jelenti, hogy két dolog/szubsztancia szükségképp *kölcsönösen* hat egymásra.<sup>2</sup> A kortárs filozófiai irodalomban azért van viszonylagos konszenzus az okság aszimmetriájának kérdésében, mert a szerzők többsége az okokat és a hatásokat eseményeknek (vagy események létre utaló tényeknek) tekinti, nem pedig szubsztanciáknak, és ez esetben valóban meggyőzőnek tűnik az a felfogás, amely szerint két esemény közül csak az egyik lehet oka a másiknak. Azonban, még ha egyetértés uralkodik is az oksági reláció aszimmetrikus voltának kérdésében, az már jelentős viták tárgya, hogy pontosan *mi határozza meg az aszimmetria irányát*.

A klasszikus értelmezés szerint a magyarázat az oksági viszony temporális jellegzetességéből adódik: az okok definíció szerint meg kell hogy előzzék okaikat.<sup>3</sup> Ez a magyarázat azonban sokakat nem elégt ki. Először is az, hogy az okok meg kell hogy előzzék okaikat, sokkal kevésbé magától értetődő követelmény, mint az okság aszimmetriája. Még ha minden további nélkül tagadjuk is annak lehetőségét, hogy a hatások megelőzzék okaikat, az mindenképp részletes magyarázatot követel, hogy miért nem lehetséges szimultán okság. A szimultán okság lehetetlenségének magyarázata viszont éppenséggel feltételezni látszik az okság aszimmetriáját. Ha ugyanis létezik oksági értelemben szimmetrikus interakció, akkor az minden valószínűség szerint szimultán is; ha két azonos töltésű részecske kölcsönösen taszítja egymást, az nyilván egy időben történik. Kétségtelen, hogy ugyanezt a jelenséget olyan temporálisan elkülönült események sorozataként is leírhatjuk, amelyek nem egyidejűek. Ennek azonban csak akkor van értelme, ha eleve feltételezzük az okság szükségszerű aszimmetriáját. Másodsor, sok filozófus véli úgy, hogy az időbeli egymásutániség legjobb magyarázata éppen az oksági reláció aszimmetriája. Mivel az események oksági sort alkotnak, az oksági viszony aszimmetriája segítségével magyarázhatjuk a temporális dimenzió sajátos (a térbeli viszonyokra nem jellemző) irányultságát.

Az aszimmetria magyarázatára számos különböző javaslat született, részben annak fényében, hogy az egyes okságelméletek számára miért jelentős az aszimmetria. A kortárs hume-iánus elméletek mindegyike valamilyen módon kapcsolatba hozza az oksági viszonyt a nomikus – tehát nem „véletlenszerű” – regularitással (Huoranszki 2001, Beebe 2006). Az „egyszerű” regularitáselmélet szerint  $C$  partikuláris esemény csak akkor lehet oka  $A$  partikuláris eseménynek, ha a  $C$  típusú események állandóan, rendszeresen vagy nagy valószínűséggel együtt

<sup>2</sup> Kant például éppen az interakció fogalma segítségével értelmezte a szimultaneitást. Vö. Kant 1787/2004. B 234.

<sup>3</sup> Hume mindkét okokról adott meghatározásában szerepel a temporális elsődlegesség. Vö. Hume 1739/1976. 241.



járnak az *A* típusú eseményekkel. Az alternatív hume-iánus elméletek szerint a nomikus regularitás valamilyen módon alá kell hogy támassza az *A* és *C* közti oksági kapcsolatot (Mackie 1973, Lewis 1983a, 1983b). Ahhoz azonban, hogy ne csak *okszági kapcsolatról*, hanem *okozásról* is beszélhessünk, meg kell tudnunk különböztetni az okokat és a hatásokat. Ezért elengedhetetlen, hogy az ok és okozat közötti viszonyt aszimmetrikusnak tekintsük.

De vajon magyarázhatja-e a pusztá együttjárás bármilyen módon is az aszimmetriát? Egyes elméletek szerint igen. Két egymástól logikailag független típusba tartozó esemény rendszeres együttjárása (pontosabban: feltételes valószínűségekkel mérhető korrelációja) ugyan nem elég ahhoz, hogy megértsük az aszimmetriát, de egy harmadik eseményhez való, ugyancsak a regularitáson alapuló viszonyuk talán segíthet ebben. Az alapötlet szerint *C* csak akkor lehet oka *A*-nak – és nem fordítva –, ha bármely harmadik *B* eseménytípus, amely *C* eseménytípussal együtt jár, együtt jár *A* eseménytípussal is (Hausman 1998. 70).<sup>4</sup> Az elmélet azon a megfigyelésen alapul, hogy a hatások bekövetkeztének több, egymástól független feltétele is van. Ezért mindig van olyan eseménytípus, amely együtt jár(hat) *A* megjelenésével, de független *C*-től. Kérdés, hogy ez így van-e. Számunkra azonban fontosabb az elmélet egy másik feltevése: minden eseménytípus, amely együtt jár – feltételes valószínűségek segítségével mérhető módon korrelál – az okkal, ugyanebben az értelemben együtt kell hogy járjon az okozattal is. Ez csak akkor lehet így, ha az okság *tranzitív* reláció, ellenkező esetben az okok okai és hatásai közt nem áll fönn szükségképpen korreláció. Az okság aszimmetriájának e magyarázata tehát megköveteli a tranzitivitást.

A kortárs hume-iánus indíttatású elméletek többsége azonban másképp kívánja magyarázni az oksági viszony aszimmetriáját. Küindulópontjuk bizonyos fizikai folyamatok lényegileg aszimmetrikus jellege. A vízbe dobott kő hullámzást okoz, a hullámok koncentrikus körökben távolodnak attól a ponttól, ahol a kő a vízbe ért. Olyan hullámjelenséget viszont nem ismerünk, amelynek során hullámok koncentrikus körökben gyűrűznének egy középpont felé (vö. Popper 1956). A legismertebb és a leggyakrabban használt példa az aszimmetrikus fizikai törvényekre a termodinamika második törvénye. Számos filozófus próbálkozik azzal, hogy az okság aszimmetriáját az entrópia szükségszerű növekedésével magyarázza. Ezen elméletek részletei igen bonyolultak, és valamennyinek elletne vethető, hogy az okság aszimmetriáját kontingens tények (bizonyos határfeltételek fennállása) segítségével képesek csak magyarázni (Elga 2000). Kérdés, hogy egy olyan univerzumban, amelyben ezek a feltételek nem állnak fenn, vajon nem kellene-e az okságot szimmetrikus relációnak tartanunk.

<sup>4</sup> Hausman maga ezt az elméletet a hume-iánus elméletek alternatívájának tekinti, az általam vizsgált kérdés szempontjából azonban ettől az önértelmezéstől eltekinthetünk. A hasonlóság kontextusfüggő, és Hausman elmélete számos, a jelenlegi tanulmány tárgya szempontjából releváns tekintetben rokon a hume-iánus elképzelésekkel.

Ez a kérdés egyesek számára talán túlzottan „metafizikainak” hat, valójában azonban egyáltalán nem kell távoli lehetséges világokról fantáziálnunk ahhoz, hogy a relevanciáját felismerjük. Érdeemes ugyanis felhívni a figyelmet arra, hogy azok a jelenségek, amelyek az okság aszimmetriáját bizonyítani hivatottak, kizárólag makrofizikai (vagy egyáltalán nem fizikai) törvényeken alapulnak. Ezért az elemi mikrofizikai szinten az aszimmetria ily módon nem magyarázható (Price 1992). Vajon nem az következik ebből, hogy az okság, ha létezik egyáltalán, tipikusan makrofizikai jelenség? Amint később látni fogjuk, abból, hogy tagadjuk a mikrofizikai folyamatok aszimmetriáját, még nem következik, hogy oksági jellegüket is meg kell kérdőjeleznünk. Számos filozófus számára a mikrofizikai folyamatok aszimmetriájának tagadása azonban elfogadhatatlan következmény lenne. Ennek oka, hogy úgy vélik, az „idő irányát” az oksági aszimmetria magyarázza. Márpedig az időnek nem csak „makrofizikai szinten” van iránya. Ami azonban fontosabb számunkra: az idő irányultsága csak akkor magyarázható az oksági viszonyok aszimmetriájával, ha azt is feltesszük, hogy az okság tranzitív reláció, mivel a temporális relációk (előbb – utóbb – egy időben) nyilvánvaló módon tranzitívek.

Az okság aszimmetriájával kapcsolatos filozófiai viták azonban azért különösen érdekesek számunkra, mert egy a tranzitivitás szempontjából alapvető belátáshoz vezettek. Lehetséges, hogy az aszimmetria megértése érdekében fel kell adnunk azt a hume-iánus hagyományt, amely okságfogalmunkat kizárólag a megfigyeléshez köti (Price 1991; Menzies–Price 1993; Price–Weslake 2009). Eszerint a szabályszerűségek megfigyelése, valószínűségek segítségével történő mérése és az ezeken alapuló törvények önmagukban nem elégségesek az oksági (és ezáltal a temporális) relációk aszimmetriájának megértéséhez. Az okság fogalmának lényegi konnotációja, hogy az okok potenciális *eszközök* a szabad cselekvő számára arra, hogy bizonyos célokat elérjen (Gasking 1955/1987; Dummett 1964; von Wright 1971/1987; Mellor 1988; Mellor 1995; Price 1991; Price–Menzies 1993).<sup>5</sup> Az okság aszimmetriáját eszerint az magyarázza, hogy a hatások sohasem lehetnek eszközök a racionális cselekvő számára az okok előidézésére. Ennek bizonyítása azonban – többek között – olyan döntéseméleti fejtegetéseket követelne meg, amelyekre most nincs módunk kitérni.<sup>6</sup>

Összefoglalva tehát úgy tűnik, az okság tranzitivitásának kérdése nem független sem az okság reflexivitására, illetve irreflexivitására, sem pedig az okság aszimmetriájára vonatkozó kérdéstől. A pusztán megfigyelt szabályszerűséget felhasználó elméletek egy része – bár bizonyítani nem tudom, feltevésem az,

<sup>5</sup> Az elképzelés természetesen nem előzmények nélküli. A legjelentősebb előzmény talán Thomas Reid okságról és szabad cselekvésről alkotott elképzelése (Reid 1788/1969). Erről részletesebben lásd Huoranszki 2002.

<sup>6</sup> Azt kell tisztázni, hogy a cselekvések választásának racionalitása független lehet-e az előzetes oksági hipotézisektől. Egyes példák azt látszanak bizonyítani, hogy nem, ámde a példák értelmezése, mint a filozófiában általában, nyitott kérdés.

hogy mindegyike – *megköveteli a tranzitivitást*; éspedig azért, mert enélkül nem tudja magyarázni az oksági viszony aszimmetriáját. Továbbá már az aszimmetria megértése kapcsán is felmerül a perspektivikusság szerepe és okságfogalmunk cél-eszköz konnotációjának jelentősége. A fentiek tehát azt sejtetik – bár nem állítom, hogy egyértelműen bizonyítják –, hogy az okság tranzitivitásának kérdése alapvetőbb, mint az aszimmetria kérdése. A következőkben azt szeretném megmutatni, hogy miért van a cél-eszköz konnotációnak kiemelt szerepe az okság tranzitivitásának megértésében is.

### III. TRANZITIVITÁS, TÚLDETERMINÁLTSAÉG ÉS OKSÁGI KONDITIONÁLISOK

Az oksági viszony tranzitivitásának elfogadása – bár tudtommal Hume sehol sem tárgyalja a kérdést explicit módon – már a hagyományos hume-i okságelméletben is megkerülhetetlen. Hume és a regularitáselméletek általában ugyanis nem tagadják, hogy nem minden szabályszerű együttjárás bizonyítja az oksági kapcsolatot jelenlétét. További feltételek kielégülésére is szükség van, és az egyik legfontosabb ilyen feltétel, hogy az ok-okozati reláció csak olyan események között állhat fenn, amelyek térben és időben érintkeznek egymással. Úgy tűnik azonban, hogy sok esetben egymással térben és időben közvetlenül nem érintkező jelenségek közt is fennáll az oksági kapcsolat. Például a budapesti orvos meggyógyíthatja betegét, ha felírja a megfelelő orvosságot, még akkor is, ha a beteg jóval később Ausztráliában gyógyul csak meg.

Miként lehetséges ilyen távolba ható okság? A választ az *okági lánc* fogalmának bevezetése nyújtja. Az egymástól térben és időben elkülönült eseményeket azon az alapon tekinthetjük oksági kapcsolatban állónak, hogy azokat egy páronként térben és időben érintkező oksági lánc köti össze. Ily módon nem kell feltételeznünk „távolba ható oksági erőt”. Viszont nyilvánvalóan fel kell tételeznünk az okság tranzitivitását; ellenkező esetben az *okági közvetítés* lehetetlenné válik. Fel kell tehát tételeznünk, hogy ha *C* oka *B*-nek, és *B* oka *A*-nak, akkor *C* egyben *A* oka is. Ebből azonban egy további fontos probléma adódik: vajon minden oksági lánc tranzitív-e? Igaz-e, hogy minden ok szükségképpen oka hatásai valamennyi következményének is?

A kortárs tranzitivitással kapcsolatos vitákat azonban elsősorban nem a „távolba hatás” lehetőségének magyarázata motiválja, hanem az úgynevezett oksági túldetermináltság problémája. Az oksági túldetermináltság minden hume-iánus elmélet számára problémát jelent, de különösen fogas kérdés a hume-iánus okságelméletek napjainkban legnépszerűbb változata, az okság kontrafaktuális kondicionálisok segítségével történő elemzése kapcsán. Ez utóbbi lényege, hogy az oksági szekvenciákat az különbözteti meg az egymással oksági kapcsolatban nem álló eseménysoroktól, hogy az utóbbiak esetében hamis, az előbbi-

ekre viszont igaz, hogy a korábbi esemény bekövetkezése nélkül az utóbbi sem *történt volna* meg. Tagadhatatlan, hogy a legtöbb olyan esetben, amikor oksági viszonyt feltételezünk két esemény között, egy ehhez hasonló kontrafaktuális állítás igazságát is elfogadjuk. További fontos kérdés azonban, hogy, először is, megérthetjük-e az oksági viszonyt pusztán a kontrafaktuálisok igazságfeltételeinek elemzése segítségével; másodszer pedig, hogy van-e valami mélyebb magyarázata annak, hogy az oksági viszonyt kifejező állítások és a kontrafaktuális állítások ilyen szorosan kapcsolódnak egymáshoz.

A problémát, mint említettem, az oksági túldetermináltság lehetősége jelenti. Az oksági túldetermináltságnak számos formája van, és – éppen a probléma „akut” volta miatt – a kortárs oksággal kapcsolatos filozófiai irodalom példák sokaságával szórakoztat (vagy untat: kinek-kinek ízlése szerint) bennünket. A következőkben azonban a túldetermináltság három típusa lesz csak érdekes számunkra. Az első, amit röviden fogok tárgyalni, az aktuális túldetermináltság. A másik a potenciális túldetermináltság két esete. Utóbbiakat azért kell részletesebben tárgyalnunk, mert ennek segítségével érthetjük meg az okság tranzitivitásának jelentőségét az okság kontrafaktuális elemzésében.

Képzelnék el – vagy inkább ne képzelnék el, csak tegyük fel –, hogy egy kivégzőosztag tagjai egyszerre találják el a szerencsétlen elítéltet. Tegyük fel továbbá, hogy szinte valamennyi lövés eltalálta az áldozatot, és hogy bármelyik önmagában is elégséges lett volna a halálához. Melyik lövés ölte meg? (Az ölés természetesen oksági fogalom, a halál szándékos *okozását* jelenti.) A probléma az, hogy az egyszerű kondicionális elemzés alapján *egyik sem*. Ha ugyanis bármelyik lövés az áldozat halálát okozta volna, akkor igaznak kellene lennie, hogy nélküle az áldozat nem halt volna meg. Ez azonban az aktuális túldetermináltság esetében nem igaz. Következésképp az egyszerű kontrafaktuális elemzés nem alkalmazható az oksági túldetermináltság eseteiben.

David Lewis – a kontrafaktuális elemzés klasszikusa – azt válaszolja erre a problémára, hogy egyetlen elméletnek sem lehet felróni, ha nem tud megoldani egy olyan kérdést, amelyre az összes többi sem tud választ adni, mivel nincs róla világos „preteoretikus” meggyőződésünk (Lewis 1983c. 194). Lewisnak ebben minden bizonnyal igaza van, de véleményem szerint ebből nem következik, hogy az aktuális túldetermináltság lehetősége ne lenne speciális ellenvetés az okság kontrafaktuális elemzésének kontextusában. Igaz, a túldetermináltság jelenségét minden elméletnek meg kell magyaráznia, és ez – az említett oknál fogva – minden elmélet számára problémát jelent. A kontrafaktuális elemzés keretében azonban az aktuális túldetermináltság nem egyszerűen probléma, hanem kifejezett paradoxon. Az elemzésből ugyanis az következik, hogy külön-külön egyik megelőző esemény sem volt oka a hatásnak. Ebből viszont az látszik következni, hogy vagy együttesen sem lehettek annak okai, vagy mindegyik csak szükséges feltétele, de – adott körülmények között – nem elégséges oka a hatás bekövetkeztének. Utóbbi esetben értelmetlen túldetermináltságról beszélni, az

előző esetben pedig a túldeterminált hatások ok nélkül következnének be. Ezzel szemben nem minden okságelmélet számára jelent önmagában nehézséget az, hogy a túldetermináltság esetében minden egyes okot *mint* a hatás egy elégséges okát azonosítsák.

Egy másik lehetséges megoldás szerint finomabb elemzés kimutathatja, hogy valójában aktuális túldetermináltság nem létezik. A kezdetben meggyőzőnek tűnő példákról az események „finomra hangolásával” talán kiderülhet, hogy a hatás egyik esetben sem volt aktuálisan túldeterminált. Fenti példánknál maradvá, lehetséges, hogy mindig volt egy olyan golyó, amelynek hatása néhány nanoszekundum eltéréssel ténylegesen megelőzte a másik – ekkor már csak potenciális – hatását. A magam részéről úgy gondolom, hogy az események „finomra hangolásával” nem oldható meg az aktuális túldetermináltság kérdése, ami szerintem logikailag és fizikailag is lehetséges. Miután azonban jelen írás tárgya az okság tranzitivitása, nem pedig a túldetermináltság, az ezzel kapcsolatos fejtegetéseket más alkalomra kell hagynom.

A „finomra hangolás lehetősége” azonban átvezet bennünket a számunkra fontosabb kérdéshez, a potenciális túldetermináltság eseteihez. Tegyük fel, hogy valamilyen logikai vagy fizikai okból kifolyólag az aktuális túldeterminálás lehetetlen. Ettől még lehetséges az aktuális *okági megelőzés* („pre-emption”), ami viszont potenciális túldetermináltságot von maga után. Mit jelent ez? Az egyszerűség kedvéért tegyük fel, hogy a kivégzőosztag mindössze két tagból áll, egy tapasztaltabb és egy kevésbé tapasztalt céllövőből. Utóbbi tüzel először. Ha nem talál, akkor következik a másik, „biztosabb kezű”. Ha már az első lövés talál, akkor az mint a halál oka nemesak időben, de okságilag is „megelőzi” a másodikat.<sup>7</sup> De nem igaz, hogy az első lövés nélkül az áldozat nem halt volna meg.

És éppen ezen a ponton válik relevánssá az okság tranzitivitásának kérdése. Lewisnak ugyanis van egy klasszikus, sokak által elfogadott – és még többek által vitatott – megoldási javaslata a megelőző okság problémájára. A javaslat abból indul ki, hogy a kontrafaktuális függés, bár elégséges, nem szükséges az oksági viszony fennállásához. Ennek oka az, hogy a kontrafaktuális függés nem tranzitív. Igaz lehet, hogy ha tegnap este a lakásom közelében találtam volna parkolóhelyet, szerencés lettem volna. Az is igaz, hogy ha szerencés lettem volna, ötösöm lett volna a lottón. De az már aligha igaz, hogy ha tegnap este találtam volna parkolóhelyet a lakásom közelében, akkor ötösöm lett volna a lottón.<sup>8</sup> Következésképp, még ha igaz is, hogy *B* esemény nem történt volna meg,

<sup>7</sup>Érdemes megjegyezni, hogy a kettő nem feltétlenül esik egybe. Képzeljük el, hogy valaki a kivégzés előtt halálos mérget tett az áldozat italába, amely azonban csak nem sokkal a lövés eldördülése után fejtette volna ki mortális hatását. Ebben az esetben a lövés csak okságilag, de nem időben előzte volna meg az áldozat halálának potenciális okát (amennyiben potenciális oknak ebben az esetben a mérgezett ital elfogyasztását tekintjük).

<sup>8</sup>Nem minden filozófus ért egyet abban, hogy a kontrafaktuális kondicionálisok intranzitívak-e vagy sem. Az ellentétes nézetről lásd Lowe 2002. 142–154. Bizonyos – itt nem tárgyal-

ha  $C$  esemény nem történt volna meg, és  $A$  esemény nem történt volna meg, ha  $B$  esemény nem történt volna meg, ebből nem következik szükségképpen, hogy  $A$  esemény nem történt volna meg, ha  $C$  esemény nem történt volna meg. Az oksági reláció viszont, állítja Lewis, tranzitív kell, hogy legyen. Ezért amikor azt állítjuk, hogy  $C$  oka  $A$ -nak, mindig feltételezzük, hogy létezik egy olyan *tranzitív oksági lánc*, amely összeköti  $C$  eseményt  $A$  eseménnyel, és amelynek minden *szomszédos elemét* kontrafaktuális függés „köti össze”.

Miként segít ez a potenciális oksági túldetermináltság magyarázatában? A feladat ebben az esetben az, hogy valamilyen módon megkülönböztessük az aktuális okot a pusztán potenciálistól. A probléma abból adódik, hogy az aktuális okról sem igaz, hogy ha nem lett volna jelen, akkor a hatás nem következik be; mégpedig éppen a „megelőzött” ok jelenléte miatt. Lewis szerint azonban az okokat és hatásaikat összekötő tranzitív oksági lánc feltételezése továbbra is lehetővé teszi, hogy az okságot a kontrafaktuálisok segítségével értsük meg. Az aktuális ok esetében ugyanis az oksági lánc minden egyes tagjára igaz lesz, hogy ha az előző nem történt volna meg, a következő sem lett volna jelen. A potenciális ok esetében azonban ez nem áll: nincs olyan oksági lánc, amely a potenciális okot, kontrafaktuális függésben álló események sorozata révén, összekötné az okozattal.

Lewis megoldási javaslata, mint csaknem minden, amit Lewis mond – többek között épp ezért olyan nagy filozófus –, erős intuitív alapokon áll. De nem támadhatatlan. Az egyik ellenérv a megoldással szemben az, hogy lehetséges olyan oksági megelőzés is, amelyben egy oksági lánc utolsó tagja előzi meg a másikat.<sup>9</sup> Ha például úgy módosítjuk eredeti, aktuális túldetermináltságra vonatkozó példánkat, hogy az osztag mindkét tagja pontosan a célra tartva tüzel, csak egyikük egy pillanattal korábban húzza meg a ravaszt, mint a másik, akkor az egyetlen hiányzó elem a potenciális oksági láncban maga a hatás. Az egymással kontrafaktuális függésben álló események láncolatának hiánya ebben az esetben nem magyarázza, hogy miért a korábban tüzelő fegyver okozta az áldozat halálát. Az ilyen esetekben ugyanis nemcsak az a kontrafaktuális hamis, hogy ha az osztag kicsit korábban tüzelő tagja nem tüzelt volna, az áldozat nem halt volna meg, hanem az is, hogy nem létezik olyan oksági lánc, amely összekötötte volna a később induló eseménysort az áldozat halálával. Lewis ezt az ellenvetést, úgy tűnik, élete vége felé elfogadta, és emiatt módosította – bár teljes egészében nem vetette el – az okság kontrafaktuális elemzését (Lewis 2004a).

ható – logikai okoknál fogva azonban a tranzitivitást védelmezők álláspontja nem plauzibilis, és még ha az is lenne, mint látni fogjuk, csak akkor lenne védhető, ha eleve lemondunk az okság kontrafaktuálisok segítségével történő értelmezéséről. A kontrafaktuálisok tranzitivitásának problémájáról vö. Bennett 2003. 159–168.

<sup>9</sup> Ezt hívják manapság a „késői megelőzés” (*late pre-emption*) problémájának, vö. Lewis 1983c, Menzies 1989 és 1996, Shaffer 2004.

## IV. TRANZITIVITÁS: ELLENPÉLDÁK ÉS ÉRVEK A PÉLDÁK ELLEN

Ezen a ponton azonban nem kívánom tovább követni a lewisi elmélet fejlődését, mert céloom most csak annak bemutatása volt, hogy a potenciális megelőző okok jelenléte csak akkor egyeztethető össze az okság kontrafaktuális elemzésével, ha feltételezzük, hogy az oksági lánc tranzitív. Persze, ha ezt csak azzal tudjuk igazolni, hogy a tranzitivitást fel kell tételeznünk egy amúgy is vitatott elmélet védelmében, érvelésünk nem lesz túl meggyőző. Lewis és sokan mások azonban úgy vélik, hogy általában, és minden speciális okságról alkotott elmélet megelőzően, az okságot tranzitív viszonyoknak tartjuk. Következésképp, ha minden látszólagos ellenpéldáról meg tudjuk mutatni, hogy értelmezhető oly módon is, hogy ne legyen cáfolata a tranzitivitásnak, akkor ez elégséges annak bizonyítására, hogy az okság valóban tranzitív.

A következőkben röviden szeretném illusztrálni, hogyan is zajlik az okság tranzitivitását elfogadók és elvetők közti „játsszma”. Ezt azonban nem azért teszem, hogy magam is beszálljak ebbe a esőrtébe. Később – a tanulmány befejező szakaszában – nem rejtem majd véka alá, hogy van véleményem arról, melyik oldalt érdemes támogatni. Azonban ahhoz, hogy ezt a véleményt kicsit is meggyőzően tudjam képviselni, mindenekelőtt azt kell megmutatnom, hogy az a keret, amelyben manapság a viták többsége zajlik, nem fog hozzásegíteni bennünket az okság tranzitivitásával kapcsolatos probléma jobb megértéséhez. Ennek pedig az oka, hogy az okság tranzitivitása – azokban az esetben, amikor helyes feltenni, hogy az oksági lánc valóban tranzitív – nem előfeltétele vagy korlátja az okság helyes filozófiai elméletének, hanem annak *következménye*.

Az oksági viszony tranzitivitása ellen felhozott példák egyik csoportja az, amikor egy oksági lánc utolsó tagja kontrafaktuálisan függ ugyan egy megelőző tagtól, első látásra azonban úgy tűnik, hogy a függés nem oksági. Kennedy elnök születését hosszú – bár a kellenél rövidebb – oksági lánc köti össze halálával, de talán nem állítanánk, hogy születése éppúgy oka volt halálának, mint az ellene elkövetett merénylet. Még akkor sem, ha igaz: anélkül, hogy megszületett, nem halhatott volna meg. A kontrafaktuális – és általában a hume-iánus – elméletek védelmezői azonban úgy érvelnek, hogy egy olyan elmélet következményeit, amely a legtöbb esetben helyesen magyarázza az oksági viszony természetét, még akkor is el kell fogadnunk, ha ez ellentmond az elmélet elfogadását megelőző intuíciónknak. Valakinek a születése igenis oka az illető halálának, még ha ezt első hallásra furesának találjuk is. Pontosan úgy, ahogyan az ősröbbanás oka valamennyi későbbi fizikai eseménynek (Hall 2004. 228). Az, hogy valami ok, még nem feltétlenül teszi érdekessé vagy akár informatívá.

Olyan példákat is találunk azonban az okság tranzitivitásával szemben, amelyek már a kontrafaktuális függésnek is ellentmondanak. Képzeljük el a következő esetet. Az egyik hegymászó észreveszi, hogy egy kisebb sziklatömb gurul a társa felé. A társa elé ugrik, és megakadályozza, hogy a szikla eltalálja. Utóbbi

folytatja útját, mintha mi sem történt volna. Ebben a példában az események olyan láncot alkotnak, melynek minden tagja kontrafaktuálisan függ a megelőző eseménytől. Ha a sziklatömb nem gurult volna lefelé, a hegymászó nem ugrott volna a társa elé, hogy megóvja. Ha a hegymászó nem ugrott volna a társa elé, az – tegyük föl – nem folytathatta volna sértetlenül az útját. Az események sorát oksági lánc kapcsolja össze. Ám furcsának tűnik azt állítani, hogy a sikeres mászásnak oka lett volna a hegymászó felé guruló kő. Először is, a kontrafaktuális függés aligha áll fenn: nem áll, hogy a hegymászó nem folytatta volna a mászást a feléje guruló szikla nélkül. Mint azonban a potenciális megelőzés kapcsán láttuk, az okság Lewis-féle kontrafaktuális elemzése ezt nem is követeli meg. Megköveteli azonban az oksági lánc tranzitivitását, és az ilyen és ehhez hasonló példákban éppen ez okozza a problémát. Mert nemcsak hogy kontrafaktuális függés nem áll fenn az oksági lánc első és utolsó tagja közt, de kérdésesnek látszik az is, hogy a leguruló kő bármilyen módon is oka lehetett a sikeres mászásnak.

Az ilyen típusú esetek tehát megkérdőjelezzik, hogy Lewis megoldása a potenciális okok kizárására sikeres. Emlékezzünk vissza: a megoldás az volt, hogy bár a pusztán megelőző okok esetében nincs kontrafaktuális függés ok és hatás közt, de van egy tranzitív oksági lánc, amely magyarázza az oksági hatást. Ámde, úgy tűnik, nem szükségszerű, hogy minden oksági lánc tranzitív. Csakhogy ha *C* esemény kontrafaktuálisan független *A* eseménytől, és az egymással kontrafaktuális függésben álló események nem alkotnak szükségképp tranzitív oksági láncot, akkor a kontrafaktuális elemzés nem tudja megmagyarázni, hogy a potenciális okokat aktuálisan megelőző okok miért is okok. Én azonban nem ezzel a kérdéssel szeretnék foglalkozni, és nem is azzal, magyarázható-e a potenciális túldetermináltság a kontrafaktuális elemzés keretein belül a tranzitivitás feltételezése nélkül.<sup>10</sup> A tranzitivitás problémájának jelentőségét ugyanis másban látom. De az, ahogyan Lewis és mások igyekeznek a tranzitivitást a fenti ellenpéldák kapcsán magyarázni, közelebb visz majd saját elképzelésem kifejtéséhez.

Lewis válasza a fenti ellenpéldákra az oksági út (*causal route*) fogalmának bevezetése (Lewis 2004a).<sup>11</sup> Az elképzelés szerint ugyanaz az esemény különböző oksági utakon keresztül is bekövetkezhet. A sikeres mászás oksági feltétele, hogy egyetlen leomló sziklatömb se vágja agyon a hegymászót. E feltétel teljesülésének egyik módja, hogy – adott körülmények között – egyetlen sziklatömb sem omlik le. De oka lehet az is, hogy jelen van a Megmentő, aki a leomló szik-

<sup>10</sup> Ahogyan például McDermott 1995 teszi.

<sup>11</sup> Ned Hall 2000 szerint az általam fentebb leírt esetre nem alkalmazható Lewis megoldása, mert az „okszági utak közti váltás” feltételezi, hogy mindkét oksági lánc (a ténylegesen megvalósuló és a „leváltott”) ugyanahhoz az eseményhez vezetett (volna). Ez a példában is használt esetekre nem áll: a sziklatömb leomlása nem vezethetett a sikeres mászáshoz. Véleményem szerint azonban Hall érvei vitathatók. Ráadásul, mint látni fogjuk, a tranzitivitás kérdése szempontjából lényegtelen, miként értelmezzük az ilyen eseteket, mivel Hall szerint az oksági reláció olyankor is tranzitív, amikor tagadjuk az események közti kontrafaktuális függést.



latömböket elhárítja az útból. Ezért az, hogy a sziklatömb leomlik-e vagy sem, okságilag meghatározza, *milyen módon* teljesül e feltétel. A leomló sziklatömb azáltal oka a sikeres mászásnak, hogy meghatározza azt az oksági (nem térbeli) utat, amelynek révén a mászás sikerül. A sziklatömb leomlása úgy működik, mint egy váltó (*switch*) a két eseménysor közt. A „váltók”, pontosabban a *váltások* pedig okai valamennyi őket követő, oksági láncot alkotó eseménynek.

De valóban úgy gondoljuk, hogy azok? Lewis megoldása a tranzitivitással kapcsolatban elvezet bennünket a kontrafaktuális elemzéssel kapcsolatos alapvető nehézséghez. Azt senki sem tagadná, hogy számos esetben – talán valamennyi tipikus esetben – van *valamilyen* kapcsolat az oksági állítások igazságfeltételei és a megfelelő kontrafaktuálisok igazsága közt. A kérdés azonban az, mi magyarázza ezt. A legjobb magyarázat – amit egy helyen Lewis maga is említ –, hogy olyan eseményekről állítjuk, hogy okok, amelyek *számítanak* (*make a difference*) a hatások létrejöttében (Lewis 1983a. 161). Ezért gondoljuk úgy, hogy a hatások nem következtek volna be, ha az okok nem történtek volna meg. Csakhogy a fenti típusú példák éppen ennek az alapvető meggyőződésnek mondanak ellent, mivel ezekben az esetekben a hatás bekövetkezett volna az állítólagos okok nélkül is. Tehát az állítólagos okok egy bizonyos értelemben nem számítanak a hatás létrejöttében. Hogy pontosan milyen értelemben, azt egy másik példával érzékeltethetjük.

Képzeld el, hogy egy szó szerint fékevesztett vonat száguld egy nagy pályaudvar felé, ahol érkezése súlyos károkat okozhat. Tegyük fel továbbá, hogy egy kisebb állomáson a pályaudvar felé vezető sínek kettévágnak, majd később a pályaudvar előtt a két sínpár újra összeér. Az első elágazásnál egy váltóór áll, aki mindezzel tisztában van. Az ő döntése határozza meg, hogy a szerelvény melyik sínpáron robot tovább a pályaudvar felé. Ez a helyzet világos módon példázza azt, amikor egy esemény több módon is bekövetkezhet, és amikor az esemény tényleges bekövetkezéséhez vezető oksági lánc valamely tagjának szerepe csak annyi, hogy meghatározza, milyen „oksági úton” következik be az esemény. A váltóór viselkedése okozza azt, hogy a vonat valamelyik sínpáron halad tovább, a vonat momentuma az adott sínpáron pedig oka annak, hogy romboló sebességgel érkezik a pályaudvarra. Mégis igen furesának tartanánk azt állítani, hogy a váltóór viselkedése bármit is számít a hatás bekövetkeztében. Az, hogy a váltóór mit tesz, valóban része egy oksági láncnak, amely a katasztrófához vezet. De ha azt állítjuk, hogy a váltóór nem felelős a katasztrófáért, azt azért tesszük, mert úgy érezzük: a katasztrófa bekövetkezése vagy elkerülése szempontjából *nem számít, mit tesz*.

Az általános konklúzió a következő: ha az okság tranzitivitását azzal próbáljuk igazolni, hogy az oksági lánc valamely tagja legalábbis meghatározta azt az „oksági utat”, amely a (végső) hatás bekövetkezéséhez vezet, akkor ellentmondásba kerülünk azzal az alapvetőnek tűnő intuícióval, amely az okság kontrafaktuális elemzésének kiindulópontjául szolgált, és amely szerint csak akkor lehet

egy esemény egy másik oka, ha az előbbi megtörténte vagy meg nem történte számít a tekintetben, hogy az utóbbi bekövetkezik-e vagy sem. Azoknak tehát, akik továbbra is kitartanak az okság kontrafaktuális értelmezése mellett, a következő alternatívák közül kell választaniuk. Vagy feladják az okság tranzitivitásának követelményét, és anélkül próbálják értelmezni a kontrafaktuális elemzését (Sartorio 2005); vagy a kontrafaktuális elemzést próbálják elválasztani attól a meggyőződéstől, hogy csak az lehet oka egy adott hatásnak, ami számít annak létrejöttében (Hall 2000).

A magam részéről azonban egészen más megoldást javasolnék: a kontrafaktuális – és általában a hume-iánus – elemzések elvetését. A kontrafaktuális elemzés elválaszthatatlan attól az intuíciónál, hogy az okok számítanak (a *make a difference* értelmében) a hatások bekövetkezése szempontjából. Véleményem szerint azonban azt, hogy oka-e valami egy másik eseménynek, nem az határozza meg, hogy megtörténte számít-e a fenti értelemben. Való igaz, a váltóőr esetében erős meggyőződésünk, hogy viselkedése nem oka a pályaudvaron történt katasztrófának. Vizsgáljunk azonban meg egy másik esetet. *X* felbérelt a feledékeny *Z*-t, hogy amikor felhívja, törjön be egy hivatalba, és tüntessen el egy számára terhelő iratot. A biztonság kedvéért azonban *X* felbérelt a hivatalban dolgozó korrupt *Y*-t is, hogy ha még ott találja másnap reggel, hozza el az iratokat. Mármint, ha minden „jól sikerül”, akkor függetlenül attól, hogy *X* hogyan dönt, az iratok másnap délben nem lesznek a helyükön. *X* viselkedése meghatározza ugyan az „okági utat”, de azt már nem befolyásolhatja, mi történik az út végén. Kérdés: az, hogy *X* telefonált vagy sem, oka volt-e ebben az esetben az iratok eltűnésének?

A válasz nem egyértelmű. Természetesen minden további nélkül állíthatjuk, hogy *X* felelősége az iratok eltűnésében abból ered, hogy ő rendezte el azokat a feltételeket, amelyek később, bármit tegyen is, az iratok eltűnéséhez vezettek. Csakhogy a kérdésünk nem az, hogy *X*-et mi alapján tartjuk erkölcsileg felelősnek, hanem hogy az a viselkedés, amely nem számít a hatás bekövetkezése szempontjából, mégis oka lehet-e annak. Kétségtelen ugyanis, hogy bármit is tesz, *X* viselkedése része lesz egy olyan okági láncnak, amely a hatás bekövetkezéséhez vezet. Amellett szeretnék érvelni, hogy akármit tegyen is, az bizony oka lesz az iratok eltűnésének. A kérdés az, mi magyarázza a különbséget a váltóőr és *X* okági szerepe között? Ha úgy gondoljuk, az okági lánc nem tranzitív az egyik esetben, milyen alapon állíthatjuk, hogy a másik esetben az?

## V. TRANZITIVITÁS ÉS OKSÁGILAG RELEVÁNS TULAJDONSÁGOK

Eddig az okság tranzitivitása kapcsán csak „példálóztunk”. A legutóbbi példa azonban azt bizonyítja, hogy talán érdemes mélyebb megfontolásokat is keresnünk arra vonatkozóan, miért is gondoljuk azt, hogy az okság tranzitív reláció. Azt szeretném megmutatni, hogy a hume-iánus elméletek a tranzitivitással kap-

csolatban *logikai okoknál fogva* csak a példálózásra hagyatkozhatnak, tehát további érvekkel a tranzitivitás ellen vagy mellett nem szolgálhatnak. És ez súlyos hiányosságnak tűnik számomra, mert az oksági viszony tranzitivitásának óriási jelentősége van nemcsak az oksággal kapcsolatos elméletekben, de számos más jelentős metafizikai probléma tárgyalása kapcsán is.

A probléma a következő. Az a kérdés, hogy három vagy több partikuláré (tárgy, esemény) egymással tranzitív viszonyban áll-e, csak a partikulárék meghatározott tulajdonságai révén értelmezhető. A tulajdonság tranzitivitása határozza meg az általa instanciált tárgyak „logikai” viszonyát. Következésképp ugyanarról a három partikuláréről egyaránt állíthatjuk, hogy egymással tranzitív viszonyban állnak, és azt is, hogy nem. Ha *C* idősebb *B*-nél, és *B* idősebb *A*-nál, akkor szükségképp *C* idősebb *A*-nál is. De ha *C* apja *B*-nek, és *B* apja *A*-nak, abból aligha következik, hogy *C* apja *A*-nak is; épp ellenkezőleg, ez lehetetlen, mivel a „*V* apja *W*-nek” intranzitív reláció. A kérdés mármost az, hogy pusztán az a tény, hogy az események oksági viszonyban állnak, elégséges-e a köztük fennálló tranzitív viszony igazolásához.

Kétségtelen, hogy vannak olyan metafizikai relációk, amelyek önmagukban elégségesek ahhoz, hogy a releváns tekintetben a partikulárék közötti logikai viszonyt meghatározzák. Ilyen például az azonosság. (Már amennyiben az azonosságot valódi relációnak tekintjük. Ez ügyben vannak kétségeim, de ezeket most félretehettük.) Ha *C* azonos *B*-vel, és *B* azonos *A*-val, akkor *C* szükségképpen azonos *A*-val is. De ilyenek a temporális relációk is, mint amilyen a „korábban történt, mint” vagy az „egy időben történt”. Az oksági relációk tranzitivitásával kapcsolatos intuíciót részben talán éppen a temporális viszonyok nyilvánvaló tranzitivitása magyarázza. Ha nem lehetséges szimultán vagy időben visszafelé ható okság, akkor az okok és hatások az *egymásutániség* tranzitív idősorán kell hogy elhelyezkedjenek. Vajon igazolhatjuk-e ezzel az oksági viszony tranzitivitását? Aligha. Előbbi példánknál maradva, az „előbb született, mint” kifejezés – nyilván az idősor tranzitivitásán alapuló – tranzitivitása nem teszi önmagában tranzitívvé az „...anya lett...” relációt. Egyetlen anyát sem tesz újra anyává az, hogy a lányának gyermeke születik. Következésképp a tranzitivitást az egymással oksági kapcsolatban álló partikulárék valamely tulajdonsága révén kellene igazolnunk.

De vajon melyik tulajdonságával? Nyilvánvaló módon az okságilag releváns tulajdonságokkal. A tranzitivitás kérdése tehát ahhoz az alapvető kérdéshez vezet bennünket, hogy melyek egy partikuláré (esemény, tárgy) kauzálisan releváns tulajdonságai. Mármost bármik is legyenek ezek, a hume-iánus elméletek egy rendkívül szigorú feltételt támasztanak velük szemben. Miután ezen elméletek szerint oksági viszony csak logikailag független események között állhat fenn, az okságilag releváns tulajdonságok nem lehetnek azok, amelyekből logikailag bármi is következik arra nézve, milyen más okságilag releváns tulajdonságokkal rendelkező partikuláréval állhatnak kauzális kapcsolatban. Ez az

egyik magyarázata annak, hogy a hagyományos hume-i elmélet a képességeket vagy diszpozíciókat nem tekintette okságilag hatékony tulajdonságoknak. Ha ugyanis valami vízben oldható, de olajban nem, akkor az nyilvánvalóan logikailag maga után vonja, hogyan fog viselkedni, ha vízbe vagy olajba teszik. Ezért aztán a hume-iánus elméletek szerint az ilyen képesség jellegű tulajdonságok nem lehetnek okságilag relevánsak.

Ha azonban az okságilag releváns tulajdonságok jellegzetessége, hogy logikailag bármely más okságilag releváns tulajdonsággal összeegyeztethetők, akkor rendkívül nehéz belátni, miként érvelhetnénk az egyes oksági láncok tranzitivitása mellett vagy ellen másként, mint például. Az alternatíva csak az lehet, hogy az okság tranzitivitását a *partikulárek okságilag releváns tulajdonságainak tranzitivitása* segítségével próbáljuk igazolni. Ezt a hume-iánus elméletek nem teszik lehetővé, ami véleményem szerint nagy hiányosságuk. Az azonban, hogy valamely oksági viszony tranzitív-e vagy sem, *nem pusztán* azoknak a diszpozícióknak vagy oksági erőknak az elemzése segítségével érthető meg, amelyek, *pace* Hume, az okságilag releváns tulajdonságokat alkotják. Mégpedig azért nem, mert ha az okságilag releváns tulajdonságokat diszpozícióknak tekintjük, akkor az okságot szinte sohasem értelmezhetjük tranzitív relációként. Az okság tranzitivitásáról alkotott erős meggyőződésünk metafizikailag sajátos magyarázatot követel.

Kiindulópontként tekintsük a hume-iánusok kedvenc (amúgy tudtommal Malebranche-tól származó) példáját, a biliárdgolyókat. Az első golyó mozgásba hozza a másodikat, a második pedig mozgásba hozza a harmadikat. Ha a „mozgásba hoz” oksági viszonyt kifejező tranzitív reláció, akkor minden további nélkül állíthatjuk, hogy az első golyó mozgásba hozta a harmadikat is. Számomra azonban a legkevésbé sem világos, hogy miért lenne ez így. Milyen okságilag releváns tulajdonságra, tehát erőre vagy diszpozícióra vonatkozhat a „mozgásba hoz” kifejezés? Lévén ez a dinamika egyik alapkérdése, aligha én fogom megválaszolni. De az a feltevésem, hogy az ezen erők segítségével azonosított okságilag releváns tulajdonságok nem támasztják alá az oksági viszony tranzitivitását.

Tekintsük a „mozgásba hozásnak” az észlelés számára legelemibb formáját, az ütközést vagy kollíziót. Ahhoz, hogy ütközés révén egy tárgy mozgásba tudjon hozni egy másikat, közvetlen érintkezésre van szükség. Bármi magyarázza is a „kollíziós képességet”, maga a képesség *nem manifestálódhat* az érintkezés nélkül. Ha azonban, mint azt felteszem, az oksági viszony éppen e képesség manifestálódásában áll, akkor ez bizonyára nem magyarázhatja az oksági lánc tranzitivitását, hiszen az érintkezés *nem* tranzitív reláció, és feltevéssünk szerint az első golyó közvetlenül nem érintkezik a harmadikkal.

A fenti példa kiterjeszthető bármely olyan „mozgásba hozó erőre”, amelyet a fizikai tárgyak vagy partikulárek okságilag releváns tulajdonságának tekintünk. Bármilyen fizikai tulajdonság legyen is, egy ilyen erő gyakorlása önmagában nem magyarázza, miként épülhet rá tranzitív oksági kapcsolat. Ennek oka az,

hogy az efféle erők gyakorlásához a képességek hordozói közt meghatározott térbeli elhelyezkedésre van szükség. Ha az elhelyezkedés olyan, hogy a képesség manifesztálódik, akkor az természetesen logikailag maga után vonja a megfelelő kapcsolatban álló partikulárék közötti oksági viszonyt. A tranzitivitás kérdése azonban ekkor nem merül fel, még akkor sem, ha a kapcsolat nem csak két partikularé között áll fenn. Például lehetséges, hogy három vagy több test kölcsönösen gravitációs hatást fejt ki egymásra. Ebben az esetben azonban nincs értelme tranzitivitásról beszélni, mivel az oksági kapcsolat közvetlen. Ha viszont három partikularé elhelyezkedése olyan, hogy az első közvetlenül nem képes a harmadik mozgásállapot-változását befolyásolni, akkor bármiképp mozogjon is a harmadik, annak nem lehet az első partikularé az oka, mivel egyetlen fizikai partikularé sem rendelkezik a másikon „átívelő” oksági erővel vagy képességgel.

Mielőtt még valaki azzal vádolna, hogy képtelenségeket állítok, szeretném világossá tenni, mi *nem* következik a fentiekből. Természetesen nem azt akarom tagadni, hogy létezik egy olyan oksági lánc, amely egy partikularé állapotától vagy változásától egy vele közvetlenül nem érintkező másik partikularé állapotának változásához vezet. Csak azt állítom, hogy az oksági lánc nem feltétlenül tranzitív, sőt a mozgástörvények esetében tipikusan nem az. Ennek ellene vehető, hogy adott körülmények között az első biliárdgolyó mozgásából és a másik két golyó elhelyezkedéséből a dinamika törvényei segítségével elvileg kiszámítható, hogy mi fog történni a harmadik golyóval. Csakhogy a kiszámíthatóságból önmagából nem következik okozás. Egy klasszikus példát alapul véve, egy zászlórúd árnyékának hosszából és a Nap állásából ki tudjuk számítani, hogy a rúd milyen magas. De ez nem jelenti azt, hogy akár a rúd árnyéka, akár a Nap oka lenne a rúd magasságának. A hatásokból bizonyos peremfeltételek fennállása mellett következtethetünk az okokra. Analóg módon, a közvetlen hatások természetének ismeretéből törvények segítségével következtethetünk az egymással oksági viszonyban nem álló eseményekre, akár a múltban, akár a jövőben történnek. A következtetés episztémikus lehetősége azonban önmagában nem elégséges ahhoz, hogy az oksági viszonyt metafizikai értelemben tranzitívnek tekintsük.

## VI. PERSPEKTIVIKUSSÁG ÉS REDUKCIO

Feltevés az – és ez valóban csak egy további bizonyítást igénylő, de a fentiek fényében, remélem, nem elfogadhatatlan feltevés –, hogy azok a fizikai erők, amelyekkel a mikrofizikai folyamatokat magyarázzuk, nem tranzitívek. Következésképp az okság tranzitivitása nem értelmezhető pusztán a mikrofizikai vagy általában véve a fizikai interakciók szintjén. Azt azonban természetesen nem állítom, hogy az oksági viszonyok sohasem tranzitívek. Állításom az, hogy a tranzitivitás magyarázata éppúgy *perspektívát* követel meg, mint az aszimmetria magyarázata. Ráadásul ez a perspektíva nagyon hasonló ahhoz, mint amit fentebb

az okság aszimmetriája kapcsán említettünk. Ennek messzemenő metafizikai következményei lehetnek. A tanulmány befejező részében azonban a lehetséges következmények közül csak azokra fogok kitérni, amelyek az okság tranzitivitásának magyarázatára vonatkoznak.

Az okság mindennapi életben és tudományban használt fogalmának egyaránt lényegi mozzanata (Hugh Mellor kifejezésével lényegi „konnotációja”), hogy az okok eszközök, amelyeket használhatunk bizonyos célok elérése érdekében. Az okok e sajátossága melleleg jól magyarázza, miként kapcsolódik az okság a kontrafaktuálisokhoz. A legtöbb esetben csak akkor tekintem  $C$ -t  $A$  okának, ha  $C$  manipulálásával meg tudnám akadályozni  $A$  bekövetkezését. Éppen ebben az értelemben mondhatjuk, hogy  $C$  bekövetkezése számít  $A$  bekövetkezése szempontjából.<sup>12</sup> A kontrafaktuális függés azonban *nem alapvető* az okság tranzitivitásának megértésében, és nem is szükséges feltétele egy tranzitív oksági lánc létének. Csak azokban az esetekben releváns, amikor a kontrafaktuális függés következik abból a tényből, hogy adott körülmények között az okok eszközök lehetnek bizonyos célok eléréséhez.

Térjünk vissza egy pillanatra ismét a három biliárdgolyó esetére. Sokan valószínűleg a következőképpen érvelnének. Azt állítottam, hogy ha az okságilag releváns tulajdonságokat képességeknek vagy diszpozícióknak tekintjük, akkor nincs megfelelő alapunk arra, hogy feltegyük: az első golyó mozgása okozta a harmadik mozgását. Azonban erős meggyőződésünk – bárki megmondhatja, aki próbálkozott már biliárdjátékkal –, hogy az első golyó mozgása igenis oka a harmadik mozgásának. Tehát egy egyszerű *modus tollens* segítségével levonhatjuk a következtetést: az okság nem képességek manifesztációja. Az eb azonban épp e ponton van elhantolva: pontosan *miért* is vagyunk annyira meggyőződve arról, hogy az első golyó mozgása oka a harmadik mozgásának, ha nem tudjuk kijelölni azt az okságilag releváns tulajdonságot, ami ezt magyarázza? Úgy vélem, azért, mert az első golyó mozgásával *képesek vagyunk vagy lehetünk elérni azt célt*, hogy a harmadik a lyukban végezze. Nem helytelen tehát azt állítani, hogy az első golyó mozgása oka a harmadik mozgásának. De csak azért nem, mert mindkettő része egy olyan oksági láncnak, amely szerepet játszik egy meghatározott cél elérésében. Azaz, az okság tranzitivitását az magyarázza, hogy bizonyos oksági láncokra *úgy tekintünk*, mint *komplex eszközökre* bizonyos célok elérése érdekében.

Ennek következménye, hogy a tranzitivitást nem annyira az okok, mint inkább a célok, azaz a hatások felől érdemes értelmezni. A célokat *valami által* érjük el. Egyszerű célokat egyszerű folyamatok által. Ha a célom az, hogy üzenetet küldjek a szomszédomnak, akkor bedobok egy levelet a postaládájába, ami által ő azt megtalálja, ami által elolvassa, ami által... Ha a célom az, hogy úrszondát küldjek a Marsra, akkor fogok egy csavarhúzózt, és bizonyos körülmé-

<sup>12</sup> Az oksági kontrafaktuálisok és a manipulálhatóság kapcsolatáról vö. Woodward 2003. 4. fejezet.

nyek között mozgatom, ami által úrszondát építek, amit útjára bocsájtok, ami által... Mindkét esetben igaz lesz, hogy létezik egy tranzitív oksági lánc az üzenet megírásától, illetve az úrszonda megtervezésétől az üzenet, illetve a szonda célba éréséig. A tranzitivitást pedig nem az oksági diszpozíciók önmagukban, hanem az „által-reláció” (*by-relation*) tranzitivitása biztosítja.<sup>13</sup> Ez a reláció azonban csak a cél-eszköz kontextusban, az elérendő célok felől értelmezhető. Ami „meggyőződésünkkel teszi” az okság tranzitivitását, az éppen az, hogy az okokat bizonyos kontextusban eszközöknek tekintjük, amelyeket egy többé vagy kevésbé komplex „által-reláció” köt össze elérni kívánt hatásaikkal.

Abból, hogy csak akkor lehet egy oksági lánc egyik elemét egy vele közvetlenül nem érintkező másik esemény okának tekinteni, ha része az események olyan sorozatának, amely eszközöknek tekinthető egy cél elérése érdekében, természetesen nem következik, hogy ennek az oksági láncban szerepet játszó partikulárék perspektívájából is értelme lenne. Nem arról van szó, hogy az első biliárdgolyó abból a célból mozog meghatározott módon, hogy a harmadikat a lyukba jutassa. Hasonló módon, nem arról van szó, hogy a napraforgó viselkedését azzal kellene magyaráznunk, hogy a növénynek magának valamilyen szándékot tulajdonítunk; vagy hogy a Fermat-elv miatt a fényről feltételeznünk kellene, hogy a legrövidebb úton igyekszik eljutni egyik pontból a másikba. Nem arról van tehát szó, hogy bizonyos esetekben úgy magyarázzuk egy partikuláré viselkedését, *mintha* szándékai volnának.<sup>14</sup> Javaslatom az, hogy egy eseményt csak akkor tekinthetünk egy vele közvetlenül nem érintkező másik esemény okának, ha egy *szabad és racionális* cselekvő azért „hozná létre”, hogy valamely hatást elérjen vele.

Továbbá a fentiekből természetesen az sem következik, hogy egy eseményt csak akkor tekinthetünk egy másik okának, ha *ténylegesen* arra használta valaki, hogy valamely célt elérjen vele. Elegendő, hogy bizonyos kontextusban értelmes dolog úgy tekinteni bizonyos típusú oksági láncokra, mint amelyek első eleme eszköz lehet egy bizonyos cél elérésére. Ebben az értelemben természetesen tranzitívnek tekinthetem a három részecske mozgása által alkotott oksági láncot is: ha az lenne a célom, hogy a harmadik részecskét mozgásba hozzam, és ezt közvetlenül csak az első mozgásával érhetném el, akkor a második részecskét az első által hoznám mozgásba, ami által mozgásba hoznám a harmadikat, stb. Csak akkor tekinthetjük azonban a folyamatot tranzitívnek, ha az része lehet egy ehhez hasonló cél-eszköz jellegű érvelésnek.

<sup>13</sup> A „by-relation” logikájának eredeti, cselekvéseméletben használt értelmezéséről lásd Goldman 1970 és Ginet 1990. A cselekvéseméleti kontextuson kívül lásd egyebek mellett Lombard 1986 és Menzies 1988.

<sup>14</sup> Javaslatom tehát nem keverendő össze Dennett „intencionális hozzáállásról” alkotott elméletével (vö. Dennett 1998). Ha valamivel, inkább Arisztotelész azon felfogásával rokon, amely szerint az okság fogalmához lényegileg hozzátartozik a valamely cél érdekében tevékenykedő ágens jelenléte (vö. Broadie 2009).

Mármost mindezek alapján miért gondoljuk úgy, hogy a szerencsétlen váltóőr nem volt oka a pályaudvaron bekövetkező katasztrófának, *X* viselkedése ellenben oka volt az iratok eltűnésének? Hiszen az utóbbi esetben is igaz, hogy adott körülmények között *X* akár telefonál, akár nem, az iratok el fognak tűnni. Ám *X* cselekedete (vagy a cselekedettől való tartózkodása) mégis oka az iratok eltűnésének, mert akárhogy cselekedjék is, az része egy olyan oksági láncnak, amelynek minden tagját az „által-reláció” köti össze abban az értelemben, hogy az adott cél eléréséhez döntése *kihagyhatatlan eleme valamelyik* oksági láncnak, amely a hatás bekövetkezéséhez vezet. A váltóőrre azonban ez nem áll. Jelenléte egyáltalán nem feltétele annak, hogy létezzék egy olyan oksági sor, amely a fék meghibásodásától vagy megrongálásától a pályaudvaron bekövetkezett katasztrófa-hoz vezet. Érzésem szerint ezért vonakodnánk attól, hogy viselkedését a katasztrófa okának tekintsük.

Végezetül egy alapvető metafizikai kérdés: vajon nem következik-e javaslatomból, hogy az okságot úgy értelmezem, mint Hume a hagyományos interpretáció szerint, azaz szubjektív módon, saját attitűdjeink kivetítéseként? Erre azt válaszolnám: igen is, meg nem is. Mindenekelőtt véleményem szerint – Hume-mal és az őt követő hagyománnyal ellentétben – minden oksági viszony olyan oksági diszpozíciók manifesztációjára épül, amelyek metafizikai szükségszerűséggel következnek a partikuláris és tulajdonságaik elrendeződéséből. Ennyiben tehát elméletem éles ellentétben áll a hume-i és általában a hume-iános felfogásokkal, és teljes mértékben „objektivista”. Másodszor – és ettől nem függetlenül – azt, hogy egy adott oksági lánc valóban hatékony eszköz lehet-e egy bizonyos cél elérése érdekében, teljesen objektív kérdésnek tekintem. Mi több – bár emellett itt már nincs módom érvelni – meggyőződésem, hogy az emergens oksági erők lehetőségét is ennek a belátásnak a segítségével magyarázhatjuk azokban az esetekben, amikor bizonyos típusú tranzitív oksági láncok változó feltételek mellett ugyanazon típusú eseményekhez vezetnek.

Ennek ellenére javaslatomnak tagadhatatlanul van egy „szubjektivista” vagy inkább projektivista vonzata, ezt azonban a magam részéről nem tekintem elvetésnek, inkább az elmélet következményének látom. Ez pedig az, hogy a tranzitív oksági láncokon szupervenialó diszpozíciók és képességek tulajdonítása „válaszfüggő”, a szó manapság bevett filozófiai értelmében. Vagyis efféle objektív diszpozíciókat és képességeket csak bizonyos mentális képességekkel és diszpozíciókkal rendelkező lények perspektívájából lehetséges tulajdonítani, következésképp nem redukálhatók mikrofizikai folyamatokra.<sup>15</sup> Ha nincs, aki értelmet adjon viselkedésének, egyetlen részecske vagy makrofizikai tárgy sem lát messzebb a közvetlen szomszédjánál.

<sup>15</sup> További kérdés, hogy mi következik ebből az idő irányultságára nézve. Ha elfogadjuk, hogy az idő irányultságát az oksági reláció természete magyarázza, lehetséges, hogy az idő irányát is csak egy sajátos perspektíva szempontjából értelmezhetjük. Ezt sem tartom elfogadhatatlannak, de jelen pillanatban nem köteleződnék el mellette. Van ugyanis egy másik, nem kevésbé plauzibilis lehetőség is, amely szerint az idő irányultsága független az okságtól.



## IRODALOM

- Beebe, Helen 2006. Does Anything Hold the Universe Together? *Synthese*. 149. 509–515.
- Beebe, Helen – Christopher Hitchcock – Peter Menzies 2009. (szerk.) *The Oxford Handbook of Causation*. Oxford, Oxford University Press.
- Bennett, Jonathan 2003. *A Philosopher's Guide to Conditionals*. Oxford, Clarendon Press.
- Broadie, Sarah 2009. The Ancient Greeks. In Helen Beebe – Christopher Hitchcock – Peter Menzies (szerk.) *The Oxford Handbook of Causation*. Oxford, Oxford University Press. 21–39.
- Cartwright, Nancy 1979. Causal Laws and Effective Strategies. *Noûs*. 13. 419–437.
- Cartwright, Nancy 1989. *Nature's Capacities and Their Measurement*. Oxford, Clarendon Press.
- Davidson, Donald 1980. *Essays on Actions and Events*. New York, Oxford University Press.
- Dennett, Daniel 1998. Az *intencionalitás filozófiája*. Szerk. Pléh Csaba. Budapest, Osiris–Gond.
- Dummett, Michael 1964. Bringing about the Past. *Philosophical Review*. 73. 338–359.
- Ehring, Douglas 1997. *Causation and Persistence. A Theory of Causation*. Oxford, Oxford University Press.
- Elga, Adam 2001. Statistical Mechanics and the Asymmetry of Counterfactual Dependence. *Philosophy of Science*. 68. 313–324.
- Gasking, Douglas 1955/1987. Okság és gyakorlati ajánlások. Ford. Csontos László. In Bertalan László (szerk.) *Magyarázat, megértés és előrejelzés*. Budapest, Tömegkommunikációs Kutatóközpont. 31–42.
- Ginet, Carl A. 1990. *On Action*. Cambridge, Cambridge University Press.
- Goldman, Alvin I. 1970. *A Theory of Human Action*. Englewood Cliffs, Prentice-Hall.
- Hall, Ned 2000. Causation and the Price of Transitivity. *Journal of Philosophy*. 97. 198–222.
- Hall, Ned 2004. Two Concepts of Causation. In John Collins – Ned Hall – Laurie Ann Paul (szerk.) *Causation and Counterfactuals*. Cambridge/MA, MIT Press. 225–276.
- Hausman, Daniel M. 1998. *Causal Asymmetries*. Cambridge, Cambridge University Press.
- Hempel, Carl 1965. *Aspects of Scientific Explanation*. New York, The Free Press.
- Hitchcock, Christopher 2001a. The Intransitivity of Causation Revealed in Equations and Graphs. *Journal of Philosophy*. 98. 273–299.
- Hitchcock, Christopher 2001b. A Tale of Two Effects. *Philosophical Review*. 110. 361–396.
- Horwich, Paul 1987. *Asymmetries in Time: Problems in the Philosophy of Science*. Cambridge/MA, MIT Press.
- Hume, David 1739/1976. *Értekezés az emberi értelemről*. Ford. Bence György. Budapest, Gondolat (újabb, javított kiadása: 2006. Budapest, Akadémiai Kiadó).
- Huoranszki Ferenc 2001. *Modern metafizika*. Budapest, Osiris.
- Huoranszki Ferenc 2002. Common Sense and the Theory of Human Behaviour. *The Philosophical Quarterly*. 52. 526–543.
- Kant, Immanuel 1787/2004. *A tiszta ész kritikája*. Ford. Kis János. Budapest, Atlantisz.
- Lewis, David 1983a. Causation. In *Philosophical Papers, Volume II*. Oxford, Oxford University Press. 159–171.
- Lewis, David 1983b. Counterfactual Dependence and Time's Arrow. In *Philosophical Papers, Volume II*. Oxford, Oxford University Press. 32–52.
- Lewis, David 1983c. Postscript to Causation. In *Philosophical Papers, Volume II*. Oxford, Oxford University Press. 172–214.
- Lewis, David 2004a. Causation as Influence. In John Collins – Ned Hall – Laurie Ann Paul (szerk.) *Causation and Counterfactuals*. Cambridge/MA, MIT Press. 75–106.
- Lewis, David 2004b. Void and Object. In John Collins – Ned Hall – Laurie Ann Paul (szerk.) *Causation and Counterfactuals*. Cambridge/MA, MIT Press. 277–290.

- Lombard, Lawrence Brian 1986. *Events: A Metaphysical Study*. London, Routledge and Kegan Paul.
- Lowe, E. Jonathan 2002. *A Survey of Metaphysics*. Oxford, Oxford University Press.
- Mackie, John L. 1974. *The Cement of the Universe*. Oxford, Clarendon Press.
- McDermott, Michael 1995. Redundant Causation. *British Journal for the Philosophy of Science*. 46. 523–544.
- Mellor, D. Hugh 1988. On Raising the Chances of Effects. In *Matters of Metaphysics*. Cambridge, Cambridge University Press. 225–234.
- Mellor, D. Hugh 1995. *The Fact of Causation*. London, Routledge.
- Menzies, Peter 1988. Against Causal Reductionism. *Mind*. 97. 551–577.
- Menzies, Peter 1989. Probabilistic Causation and Causal Processes: A Critique of Lewis. *Philosophy of Science*. 56. 642–663.
- Menzies, Peter 1996. Probabilistic Causation and the Pre-emption Problem. *Mind*. 105. 85–117.
- Menzies, Peter – Huw Price 1993. Causation as a Secondary Quality. *British Journal for the Philosophy of Science*. 44. 187–203.
- Molnar, George 2003. *Powers. A Study in Metaphysics*. Oxford, Oxford University Press.
- Paul, Laurie Ann 2004. Aspect Causation. In John Collins – Ned Hall – Laurie Ann Paul (szerk.) *Causation and Counterfactuals*. Cambridge/MA, MIT Press. 205–224.
- Popper, Karl R. 1956. The Arrow of Time. *Nature*. 177. 538.
- Price, Huw 1991. Agency and Probabilistic Causality. *British Journal for the Philosophy of Science*. 42. 157–176.
- Price, Huw 1992. Agency and Causal Asymmetry. *Mind*. 101. 501–520.
- Price, Huw – Brad Weslake. 2009. The Time-Asymmetry of Causation. In Helen Beebe – Christopher Hitchcock – Peter Menzies (szerk.) *The Oxford Handbook of Causation*. Oxford, Oxford University Press. 414–443.
- Reid, Thomas 1788/1969. *Essays on the Active Powers of the Human Mind*. Szerk. Baruch A. Brody. Cambridge/MA, MIT Press. Részletei magyarul: Értekezések az ember aktív erőiről. Ford. Fehér Ferenc. In Márkus György (vál., jegyz.) 1977. *Brit moralisták a XVIII. században*. Budapest, Gondolat. 565–674.
- Russell, Bertrand 1953/1976. Az ok fogalmáról. Ford. Márkus György. In *Miszticizmus és logika*. Budapest, Magyar Helikon. 291–337.
- Sanford, David H. 1989. *If P, then Q. Conditionals and the Foundations of Reasoning*. London – New York, Routledge.
- Sartorio, Carolina 2005. Causes as Difference Makers. *Philosophical Studies*. 123. 71–96.
- Shaffer, Jonathan 2004. Trumping Preemption. In John Collins – Ned Hall – Laurie Ann Paul (szerk.) *Causation and Counterfactuals*. Cambridge/MA, MIT Press. 59–73.
- Woodward, James 2003. *Making Things Happen*. Oxford, Oxford University Press.
- Wright, Georg Henrik von 1971/1987. Magyarázat és megértés. Ford. Csontos László. In Bertalan László (szerk.) *Magyarázat, megértés és előrejelzés*. Budapest, Tömegkommunikációs Kutatóközpont. 43–210.

Rovatunk tárgya ez alkalommal Farkas Katalin *The Subject's Point of View* című, az Oxford University Pressnél 2008-ban megjelent könyve. A kötet tartalmának összefoglalását Márton Miklós, Forrai Gábor és Tózsér János recenziói, majd a szerző válaszai követik.



*The Subject's Point of View*

## A könyv összefoglalása

A kortárs elmefilozófia egyik sokat vitatott kérdése a mentális állapotok természetével kapcsolatos, méghozzá azzal a kérdéssel, hogy vajon a mentális állapotok és tartalmak egyedül a gondolkodó alany belső állapotaitól függenek-e (ez az internalizmus álláspontja), avagy részben az alanyon kívüli körülmények konstituálják őket (ahogy azt az externalisták gondolják). A modern filozófiai hagyomány jellemzően internalista volt: Descartes szerint például elképzelhető, hogy egy gonosz démon megcsal, és valójában rajtam kívül nem létezik semmi, mentális állapotaim mégis ugyanazok, mint most. Tehát ezen állapotok természete lényege szerint nem függ a rajtam kívül álló dolgoktól. Ezzel szemben a kortárs ortodox nézet externalista, és az externalista elméletek a 20. és 21. században a karteziánus elmefelfogást ért széles körű kritikák egyik legfontosabb csoportját alkotják. Még azok a szerzők is, akik szimpatizálnak az internalista felfogással, általában elismerik, hogy legalábbis az elme egyes aspektusai externalista megközelítést igényelnek.

*The Subject's Point of View* című könyvemben egy kompromisszummentes internalista elmélet védelmét tűztem ki célul. Egy olyan elmefelfogás mellett érvelek, amely egy fontos szempontból karteziánus. Nem abban az értelemben, hogy az elme és a test dualizmusát kellene elfogadnunk, ebben a kérdésben ugyanis semleges maradok. Az általam javasolt elmélet inkább Descartes második elmélkedésének címéhez kapcsolódik: *Az emberi elme természetéről: arról, hogy jobban ismerjük, mint a testet*. Amellett érvelek ugyanis, hogy a mentális vonások megkülönböztető sajátossága az, hogy a gondolkodó alanyunk privilegizált hozzáférése van ezekhez az állapotokhoz; továbbá amellett, hogy ez a körülmény az elme externalista elméletének teljes elutasítását teszi szükségessé.

Az első fejezetben azt a tézist védelmezem, hogy a privilegizált hozzáférés a mentális vonások megkülönböztető jegye. A ránk jellemző vonásokat a következő kritérium alapján osztjuk mentális és nem mentális kategóriákba: pontosan az tartozik elmémhez, amit úgy ismerek meg, ahogyan azt senki más nem tudja megismerni. Ezzel nem azt tagadom, hogy mások is megismerhetik az én elmémet; mindössze azt, hogy ugyanolyan módon férnek hozzá elmémhez, mint én.

A második fejezetben megmutatom, hogy bizonyos ellenvetések az első fejezet tételével szemben alaptalanok. Az egyik ellenvetés szerint igaz ugyan, hogy egyes testi állapotaimat gyakran úgy ismerem meg, ahogyan mások nem, ez azonban e megismerés esetleges és nem szükségszerű vonása. Egy másik ellenvetés a tudattalan állapotokkal és az öncsalás eseteivel kapcsolatos: ezek ugyanis látszólag ellenpéldák azzal az állítással szemben, hogy én magam másképpen – és jobban – ismerem mentális állapotaimat, mint mások. Egy Freudtól kölcsönzött gondolatmenet segítségével bizonyítom, hogy a tudattalan állapotokról alkotott felfogás lényegileg parazitikus a tudatos, privilegizált módon hozzáférhető állapotok fogalmán.

A harmadik fejezetben az eddig védelmezett elmefelfogás jelentősége mellett érvelek. Az emberi élet egyik legfontosabb aspektusa a másokhoz való viszony: méghozzá azokhoz, akiket magunkhoz hasonlóan személyeknek tekintünk. Álláspontom szerint azokat tekintjük személyeknek, akik hozzánk hasonló elmével rendelkeznek, azaz hozzánk hasonló mentálisállapot-fajtákkal rendelkeznek. Azt azonban, hogy mi számít mentálisállapot-fajtának, a karteziánus kritérium határozza meg. Éppen ezért a személyekről alkotott felfogásunk alapvetően támaszkodik a karteziánus elméletre.

A negyedik fejezettel kezdődő második részben az eddigi eredményeket a kortárs internalizmus–externalizmus vitához kapcsolom. Először azt a kérdést vizsgálom meg, hogyan kell pontosan definiálni ezt a vitát. A szokásos felfogás szerint a kérdés azon múlik, hogy a mentális állapotokat az alany agyán vagy testen kívüli vagy belüli tényezők határozzák-e meg. A belső és külső közötti határvonal azonban nem az agy vagy a test körül húzható meg. Az internalista nézet lényege valójában az, hogy a mentális vonásokat egyedül az határozza meg, hogy a világ hogyan jelenik meg az alany számára – tehát az, hogy hogyan is állnak a dolgok a szubjektum nézőpontjából. Az externalizmus ennek tagadása.

Az ötödik fejezetben amellet érvelek, hogy akkor jelennek meg a dolgok ugyanúgy két alany számára, ha ugyanazokkal a belső fenomenális tulajdonságokkal rendelkeznek. Az internalista szerint az határozza meg mentális vonásaimat, ahogyan a dolgok számomra megjelennek. Az externalista szerint viszont mentális vonásaim lehetnének különbözőek akkor is, ha minden pontosan ugyanilyennek tűnne. A hatodik fejezetben megmutatom, hogy emiatt az externalizmus korlátozza a privilegizált hozzáférés körét. Az alany ugyanis csak ahhoz fér hozzá privilegizált módon, hogy a dolgok hogyan jelennek meg számára. Ez viszont azt jelenti, hogy az externalizmus ellentétbe kerül az elméről alkotott alapvető felfogásunkkal, amely a privilegizált hozzáférés kritériumán alapul. Tehát az externalizmus téves.

A fenti érvekhez hasonló megfontolások több filozófusban is kételyt ébresztettek az externalista elmélettel kapcsolatban, ám sokan gondolták úgy, hogy végül is az internalizmus tarthatatlan, mivel nem tud számot adni a mentális állapotok szemantikai vonásairól; arról, hogy ezek az állapotok referencia- és igaz-

ságfeltételekkel rendelkeznek. A mű hetedik fejezetében megmutatom, hogy ez az aggodalom alaptalan.

Anthony Kenny az elme metafizikájáról írott könyvében egy helyen azt mondja, hogy – Gilbert Ryle-hoz hasonlóan – ő is Descartes örökségét tekinti az elme természetének helyes megértését akadályozó legsúlyosabb akadálnak. Kenny nincs egyedül ezzel a véleménnyel. Munkám végső konklúziója az, hogy a 20. és 21. század filozófiájában széles körben elfogadott antikartezianus nézet lényegileg elhibázott. Descartes örökségének elutasítása sokkal problematikusabb, mint ahogy azt kritikusai felteszik: ugyanis az elméről és a személyekről alkotott alapvető és nélkülözhetetlen felfogásunk valójában a descartes-i nézeten nyugszik.

## Vonakodó karteziánizmus?

Farkas Katalin remek könyvében az elme természetéről szóló karteziánus elképzelést veszi védelmébe. E karteziánus elmefelfogás legfontosabb elemének azonban sokakkal szemben nem az anyagtalan szellemről szóló descartes-i tanítást tekinti, hanem a mentális állapotokhoz történő privilegizált hozzáférés tételét. Sőt, Farkas egyenesen úgy foglal állást, hogy a privilegizált hozzáférés tételéből – ellentétben azzal, ahogy az iskolában tanultuk – nem következik a szubsztancia-dualizmus. E tétel alátámasztása érdekében a privilegizált hozzáférés fogalmát igen kreatívan átértelmezi, meglehetősen speciális értelmet kölcsönözve neki. Tanulmányomban azonban amellet kívánok érvelni, hogy Farkas e vállalkozása sikertelen: a privilegizált hozzáférés még e speciális értelemben is maga után vonja a szubsztancia-dualizmust. Konklúzióm tehát az, hogy nem lehet csak félig megvenni az elme karteziánus modelljét, és vonakodni annak teljes egészében történő elfogadásától.

### FARKAS KRITÉRIUMA

Könyvének első fejezetében Farkas azt igyekszik kimutatni, hogy a mentálisnak tekinthető állapotok éppen azon emberi képességekből származnak, amelyek meglétében Descartes híres, a démonhipotézisen alapuló szkeptikus gondolkísérletének elvégzése után is biztosak lehetünk. A gonosz démon szisztematikusan megtéveszt bennünket, így semminek a létében nem lehetünk biztosak, ami közvetlen tapasztalatunkon vagy gondolkodásunkon kívül esik, beleértve közvetlen tapasztalataink és gondolataink okait is. Így ami megmarad bizonyosan létezőnek, az nem más, mint maga a közvetlen tapasztalat, vagyis azok az állapotok, melyekbe a tapasztalás vagy a kognitív képességek gyakorlása során kerülünk. Tehát ezek, vagyis a démon-teszt elvégzése után megmaradt állapotok számítanak mentális állapotoknak Farkas szerint.<sup>1</sup>

<sup>1</sup> Farkas 2008. 13.



Farkas számára azonban a démonhipotézis felhasználása nem azért fontos, hogy kiderítse, mely jelenségek létezésében lehetünk biztosak, és melyekében nem, hanem azért, mert a démon-teszt éppen azokat az állapotokat tartja fenn a rostán, melyekhez egyfajta sajátos, első személyű perspektívából férünk hozzá. A tapasztalat és a gondolat okai vagy tárgyai mindenki számára ugyanúgy, saját tapasztalatán és gondolkodásán keresztül tárulnak föl, ezzel szemben magához a közvetlen tapasztalathoz vagy gondolkodásához mindenki csak a saját személyes módján fér hozzá. E sajátos hozzáférés szerepét Farkas szerint a reflexió vagy introspekció tölti be. Álláspontja szerint tehát azokat a jelenségeket, amelyeket reflexió vagy introspekció útján ismerek meg, senki más nem ismerheti meg ezen a módon (18–19). Ez a megállapítás ugyanakkor nem zárja ki, hogy az általam introspekció útján megismert jelenségeket egy másik személy más módokon megismerhesse. Farkas tehát nem tagadja, hogy a reflexió által feltárt jelenségek bizonyos értelemben publikusak, pusztán annyit állít, hogy egyes publikus jelenségeket egy személy megismerhet oly módon is, ahogyan más számára ez nem lehetséges. Más ugyanis azt, amit én reflexió útján ismerek meg, nem tudja szintén *reflexió útján* megismerni – ugyanennek a megismeréséhez csak másfajta képességek és lehetőségek állnak a rendelkezésére. Mindez a reflexión vagy introspekción kívül más képességekre nem áll: amit képes vagyok észlelni, más is képes lehet észlelni; amit képes vagyok elgondolni, más is képes lehet elgondolni, amire képes vagyok visszaemlékezni, arra más is képes lehet emlékezni, és így tovább.<sup>2</sup> Mármost Farkas legfontosabb tézise szerint éppen ezek, a reflexió útján csak a reflektáló alany számára hozzáférhető jelenségek tartoznak az elméhez. A privilegizált hozzáférés tézise ebben a formában tehát azt állítja, hogy azok és csak azok a jelenségek mentálisak, amelyeket olyan módon ismerhetek meg, ahogyan senki más.

A szerző intenciói szerint továbbá az a tény, hogy a mentális állapotokhoz ilyen kitüntetett módon férünk hozzá, *nem zárja ki*, hogy e mentális állapotok bizonyos fiziológiai állapotoktól függjenek, amelyek esetében természetesen nem áll fenn a privilegizált hozzáférési mód tézise. Még hozzá azért nem, mert Farkas szerint a démonhipotézis nem egy anyagtalán szellemre szűkíti le a világot, hanem egy szubjektív nézőpontra, a vizsgálódás egy sajátos, első személyű szemszögére (18). Éppen ezért leszögezi: nyitott kérdés, vajon e nézőpontnak szüksége van-e testre vagy sem. A könyv egyik fontos tétele szerint tehát a démon-teszt alkalmazása a mentális állapotok listájának elkészítésére – szemben a szokásos „iskolás” értelmezéssel – önmagában nem kötelez el bennünket a szubsztancia-dualizmus mellett.<sup>3</sup> Ennek megfelelően Farkas több helyen is

<sup>2</sup> E tétel részletes kifejtését és védelmét lásd a 17–24. oldalakon.

<sup>3</sup> Persze ez az „iskolás” értelmezés valójában már Descartes esetében sem állja meg a helyét, ahogy azt Farkas világosan ki is mutatja. Lásd 34–35.

(lásd például 15 és 54) kifejezetten hangsúlyozza, hogy semleges kíván maradni a fizikalizmus kérdését illetően.

Ez tehát az az állítás, amellyel vitába kívánok szállni. Az alábbiakban egy ellenpélda bemutatása és értelmezése segítségével amellet szeretnék érvelni, hogy a speciális hozzáférés imént ismertetett kritériuma csak akkor tartható fenn a mentális állapotok megkülönböztető jegyeként, ha elköteleződünk a szubsztanciadualizmus mellett.

#### AZ ELLENPÉLDA ÉS ÉRTELMEZÉSE

Farkas könyvének második fejezetében vet számot az imént ismertetett kritérium ellen fölhozható érvekkel. A fejezet jórészt azokkal a lehetséges ellenvetésekkel foglalkozik, melyek azt igyekeznek kimutatni, hogy a kritérium nem lehet a mentális jelenségek listájára történő fölkerülés *szükséges* feltétele. Ezek az ellenérvek éppen ezért úgy működnek, hogy olyan nyilvánvalóan mentálisnak tekintett jelenségeket mutatnak föl, amelyek esetében nem teljesül a kritériumban megfogalmazott követelmény. Ilyenek elsősorban a nem okkurens és a tudattalan vagy pretudatos mentális állapotok. Farkas igen alaposan tárgyalja az e jelenségekre alapozott ellenvetéseket, és úgy vélem, meglehetősen plauzibilis válaszokat ad rájuk (vö. 39–50).

Érdemes azonban észrevennünk, hogy a kritérium voltaképpen két állítást fogalmaz meg: (a) az alany reflexió útján hozzáfér mentális állapotaihoz, és (b) senki más nem fér hozzá ezekhez reflexió útján. Nos, a Farkas által tárgyalt ellenérveknek van egy közös vonásuk: mind az első állítást vitatják. Azt teszik kérdésessé, hogy az alany bizonyos mentális állapotaihoz valóban hozzáfér-e reflexióval vagy introspekció útján. Egy sines azonban köztük, amely a második állítás igazságát kérdőjelezné meg, vagyis azt igyekezne kimutatni, hogy egy alany valamely mentális állapotához egy másik személy is hozzáférhet reflexió vagy introspekció útján.

Hogy egy ilyesfajta ellenpélda lehetséges, ahhoz éppen Farkas egyik érveléséből meríthetünk ötletet. Ebben egy olyan ellenvetésre válaszol, amely az imént említettekkel ellentétben nem azt igyekszik kétségbe vonni, hogy a kritérium szükséges feltétele volna a mentális állapotok listájára kerülésnek, hanem azt, hogy *elégséges* (33–39). Ez az ellenérv tehát azt igyekszik kimutatni, hogy nemcsak mentális, hanem pusztán fizikai, jelesül testi jelenségek esetében is működik a kritérium, vagyis nem csak mentális állapotokra igaz, hogy csak az alany fér hozzájuk egy meghatározott módon. Farkas válaszában úgy hárítja el az érvet, hogy megmutatja, valójában e testi jelenségek esetében nincs *elvi* akadály annak, hogy az alanyon kívül más is ugyanolyan módon férjen hozzájuk. Ezen a ponton merül föl a kérdés, miért ne lenne ugyanez lehetséges mentálisnak tekinthető állapotok esetében? A válasz megtalálása érdekében nézzük meg a részleteket!

A Farkas által tárgyalt ellenvetés Richard Rortynak egy 1970-es tanulmányából származik. Rorty ebben az írásában jó néhány, a „mentális” megkülönböztető jegyére aspiráló kritériumot számba vesz, és a legtöbbet elutasítja. Így tesz azzal a kritériummal is, amely szerint mentálisak azok a jelenségek, amelyekről „introspekció útján szerzhető információ” (Rorty 1970. 413). Ez Rorty szerint azért nem jó megkülönböztető jel, mert pusztán fizikai jelenségekről is szerzhetünk tudomást introspekció útján. Például, mivel sajnos sokszor előjött már a lumbágóm, ezért amikor fáj a derekam, ebből introspektíve tudom, hogy egy idegvégződésem becsípődött két porc közé. Igaz ugyan, hogy itt talán csak közvetett tudásról van szó, hiszen a közvetlen fájdalomérzet alapján „következtetek” arra, hogy az idegszállammal már megint megtörtént az a bizonyos dolog, ez azonban mindenképpen introspekció útján szerzett információ lesz, és – a sajnálatosan gyakori előfordulás miatt – meglehetősen megbízható is. Ugyanakkor első ránézésére úgy tűnik, csak én szerzhetek ilyen módon információt e testi jelenségről, senki más. Úgy tűnhet tehát, hogy itt egy ellenpéldát találtunk Farkas kritériumával szemben: a lumbágómról csak én tudhatok olyan módon, hogy fáj a derekam, miközben az idegbeesípődés nyilvánvalóan nem mentális jelenség.

Rorty ellenvetésével szemben azonban Farkas plauzibilisen érvel amellett, hogy fogalmilag nem kizárt olyan eset, amikor az én testem valamely pusztán fizikai állapotáról egy másik ember idegi információt szerez – mondjuk az agyának megfelelő részből kiinduló idegvégzódések kvázi „rá vannak kötve” az én szóban forgó testrészemre. Egy ilyen, kétségekívül lehetséges esetben mindketten pontosan ugyanolyan módon tudunk e pusztán fizikai jelenségről. Mindez azonban Farkas szerint az elme jelenségeivel kapcsolatban nem lehetséges, éppen ezért kitart a „különleges hozzáférés” kritériuma mellett. Álláspontja szerint tehát – miként ezt explicit módon ki is mondja (24) – a testi érzet, a testi tudatosság (*bodily awareness*) nem ugyanaz a megismerési mód, mint az introspektív tudás. Míg az előbbi, mint láttuk, nem szükségképpen aszimmetrikus, addig az utóbbi Farkas szerint az.

Farkas e tényt a *perspektivikusság* metaforájával érzékelteti: saját mentális állapotaihoz – és a mentális képességek által feltáruló környezetéhez – az ember (vagy más élőlény) sajátos perspektívával rendelkezik (vö. 30–31). E perspektivikusságnak továbbá nyilvánvalóan *szükségszerűnek* kell lennie, hiszen – mint láttuk – az pusztán esetlegesség, hogy az idegsebészlet jelenlegi állapota mellett történetesen csak én ismerhetem meg a lumbágómat olyan módon, hogy fáj a derekam. Arról nem is beszélve, hogy ha elegendő lenne a kontingens értelemben vett perspektivikusság, akkor a szó eredeti, térbeli értelmében vett perspektivikus tények is mentálisak lennének, mint például egy szobornak a tér egy pontjáról megfigyelhető látványa. Ahhoz, hogy a kritérium megkülönböztető jegyként működjön, az kell, hogy a mentális jelenségek esetében *lehetetlen* legyen, hogy két személy ugyanabból a perspektívából ismerje meg a jelenséget.

Vagyis a Rorty által felhozott testi érzetek példája pusztán praktikus és nem konceptuális kihívást jelent Farkas kritériuma számára (vö. 35).

Véleményem szerint azonban éppen a szükségszerűség e követelménye miatt nem működik Farkas kritériuma. Ahogyan pusztán technikai probléma az én idegbecsípődésem más ember agyába történő „bedrótozása”, éppúgy pusztán technikai, és nem elvi nehézséget okoz annak elképzelése, hogy a derékfájásomért felelős agyterület(ek)et pontosan úgy „drótozzák be” egy másik személy agyába, ahogyan az én agyamban e bedrótozás megvalósul. Képzeljük el például, hogy a távoli jövőben az idegsebészet képes lesz arra, hogy a született vagy balesetben szerzett agykárosodással élő emberek számára olyan implantátumot gyártson, amely tökéletesen alkalmas a sérült agyterület funkciójának ellátására. Ehhez mindössze az implantátumnak az alany agyába történő megfelelő „bedrótozásáról” kell gondoskodni, ami e fejlett idegsebészet számára nem jelent akadályt. Tegyük fel, hogy én – pontosabban a távoli jövőben élő alteregóm – úgy születtem, hogy a derékfájdalom érzetért felelős agyterületem olyan mértékben sérült, hogy egyáltalán nem érzek fájdalmat akkor, amikor az idegbecsípődés megtörténik. A megoldás azonban nagyon egyszerű: egy rutinműtét során agyamba ültetik a megfelelő implantátumot. Képzeljük el azonban, hogy e távoli jövőben, bár az idegsebészet nagyon fejlett, az ország gazdasági helyzetéről közel sem mondhatunk ilyen jókat. Éppen ezért – az egészségbiztosítás szűkös forrásai miatt – egy hasonló problémával küzdő társammal együtt, közösen kapunk egy ilyen implantátumot, melyet mindkettőnk agyába „bedrótoznak”, és így közösen használjuk azt.<sup>4</sup> Sőt, akár egy olyan scenáriót is elképzelhetünk, amelyben kombináljuk az iménti példát Farkas már említett gondolatkísérletével, vagyis csak az én porckorongjaim felől jövő idegeket „kötnék be” egy implantátumba, amit aztán mind az én, mind egy másik ember agyába „bedrótoznának”. Ez a másik ember lehet mondjuk egy orvos, aki meg akar győződni róla, hogy az implantátum jól működik-e – ennek érdekében tehát kipróbálja, tényleg érzem-e a derékfájdalmat.

Nos, értelmezésem szerint ezekben az elképzelt esetekben mindkét személy *ugyanolyan módon* ismerne meg egy mentális állapotot. Farkas kritériuma tehát nem lehet szükséges feltétele annak, hogy egy esemény vagy állapot mentálisnak számítsion. Sőt, álláspontom szerint fenomenológiailag sem nagyon lehetne megkülönböztetni a két ember által átélt élményt, hiszen mindketten úgy éreznék, az ő derekuk fáj, méghozzá pontosan ugyanúgy. Farkas ugyan többször hangoztatja, hogy a mentális állapot – mondjuk egy fájdalom – azért nem lehet ugyanaz két, akár a releváns értelemben „összedrótozott” embernél, mert a másik ember a fájdalmat mint az ő fájdalmát éli át, nem pedig mint az enyémet,

<sup>4</sup> Akár azt is elképzelhetjük, hogy a „bedrótozás” ebben az esetben levegőben terjedő elektromágneses hullámok segítségével történik, úgyhogy a szó hagyományos értelmében nem leszünk fizikailag összekötve.

és éppen ezért az általa érzett fájdalom is az övé, nem pedig az enyém (például 37). Ezzel szemben úgy vélem, éppen a mentális állapot, a fájdalom átélésének fenomenológiája az, ami végképp nem tesz különbséget a két állapot között, hiszen az mindkét személy számára mint a *saját* fájdalma jelenik meg. Farkas metaforáját fölhasználva tehát azt mondhatjuk, hogy éppen ezért ugyanabból a perspektívából tárul föl számukra mind a fájdalom, mind – ha annak testi kiindulópontja is közös – annak oka.

IHa tehát igaz, hogy ahhoz, hogy nekem fájjon a derekam, elegendő, ha ez a bizonyos esemény lejátszódik azon a bizonyos agyterületen vagy implantátumban, ami megfelelő módon „be van drótozva” az agyam többi része közé, akkor annak is igaznak kell lennie, hogy a másik ember *pontosan ugyanazt* a derékfájást érzi, mint én. Annak a közhelynek, hogy két ember nem érezheti ugyanazt a fájdalmat, éppen annyi esélye van egy fejlett idegsebészet fényében, mint annak, hogy két embernek nem lehet ugyanaz az ujjlenyomata – egy fejlett plasztikai sebészet fényében.

Az persze igaz, hogy továbbra is marad némi aszimmetria a két „összedróttozott” személy fájdalomészlelése között, mégpedig az, hogy mindketten közvetlenül csak azt fogják tudni, hogy *nekik* derékfájdalmuk van, arról legfeljebb közvetett empirikus úton értesülnek, hogy a másik személynek is. Azonban e különbség ismeretelméleti, és nem metafizikai különbség: nem érinti azt a tényt, hogy ugyanazt a jelenséget ismerik meg ugyanolyan módon. Episztemikus helyzetükben fennálló különbségük ugyanis *nem* a közösen érzett fájdalomérzetükre vonatkozik, hanem egymás pszichológiai működésére. A perspektivikusság eredeti térbeli jelentésében mindez nyilvánvaló: abból, hogy én csak magamról tudhatom közvetlenül, hogy egy bizonyos szemszögből szemlélek egy szobrot, nem következik, hogy más ne szemlélhetné ugyanabból a szemszögből *ugyanazt* a szobrot.

## AZ ÉRTELMEZÉS VÉDELME

Az iménti szakasz konklúziója szerint egy mentális állapothoz, például egy fájdalomhoz megfelelő neurális elrendezés esetén több személy is ugyanolyan módon képes hozzáférni. Ennek cáfolatához azt kellene kimutatnunk, hogy még a fenti gondolat kísérlet esetén is az a helyzet, hogy a két személy két, numerikusan különböző fájdalommal rendelkezik. Én ellenben amellet szeretnék érvelni, hogy ez csak akkor lehetséges, ha elfogadjuk a szubsztancia-dualizmust, azaz tagadjuk akár a legenyhébb értelemben vett fizikalizmust is.

Persze érdemes először tisztáznunk, mi is a fizikalizmus. Farkas maga ehhez nem ad sok segítséget, könyvében ugyanis nemcsak a fizikalizmus érvényesége ügyében, hanem abban a kérdésben sem akar állást foglalni, hogy hogyan kell értenünk a fizikalizmust. Nos, a legtöbb mai monográfia és kézikönyv a fizi-

kalizmust a logikai ráépüléssel definiálja.<sup>5</sup> A logikai ráépülés fogalma a mentális jelenségeknek a fizikaiaktól való olyan függését hivatott kifejezni, amely szerint *lehetetlen*, hogy mentális különbség álljon elő fizikai különbség nélkül. A logikai ráépülés szélső esetét képviseli a pszichofizikai azonosság tézise, ugyanakkor önmagában e fogalom a mentálisnak a fizikaitól való kevésbé szoros függését is megengedi. Ami viszont fontos, hogy a logikai ráépülés egy olyan függést fogalmaz meg, amely erősebb a mentális és a fizikai pusztai oksági kapcsolatánál, hiszen ez utóbbi – a legtöbb értelmezés szerint – pusztán kontingens viszonyt fejez ki.

Mindezt figyelembe véve, ráépülhet-e mármost a fenti ellenpéldában szereplő pusztán fizikai állapot (az agyterület vagy az implantátum állapotának) valamely tulajdonságára két, numerikusan különböző mentális állapot vagy tulajdonság? Nos, bár logikailag, a ráépülés fogalmából kiindulva nem látok olyan érvet, ami ezt kizárná, mindazonáltal némi megfontolás után olyannyira implauzibilisnek tűnik számomra, hogy úgy vélem, ezt az elképzelést nyugodtan elvethetjük.

Először is szögezzük le, hogy a fenti esetben a mentális és a fizikai állapotok azonosságáról nem lehet szó, hiszen ha ragaszkodunk hozzá, hogy két mentális állapot van jelen, akkor az azonosság tranzitivitása miatt a különbözőnek vélt mentális állapotokról is kiderülne, hogy azonosak. Ami az azonosságnál gyengébb ráépülési viszonyt illeti, abban persze semmi szokatlan nincs, hogy általában véve egy tulajdonságra több másik épül rá. Így gondolunk például arra az esetre, amikor az alma kémiai szerkezete egyaránt meghatározza annak állagát és ízét. Esetünkben azonban ennél sokkal furcsább lehetőséget kell elképzelnünk. Azt tudniillik, hogy két, egymástól kvalitatíve semmiben sem eltérő, *pusztán numerikusan* különböző mentális állapot épül rá ugyanazon fizikai állapot valamely tulajdonságára. Nyilvánvaló ugyanis, hogy nem lehet minőségi különbség a feltelezett két mentális állapotban, hiszen ekkor olyan különbség lenne a mentálisban, amit nem magyarázna valamilyen fizikai különbség, ami ellentmond a logikai ráépülés fenti definíciójának.

Mármost ez a lehetőség teljesen abszurd. Ha ugyanis ezt megengednénk, akkor azt is meg kellene engednünk, hogy a normális esetben az én fájdalomközpontként funkcionáló agyterületem állapotára vagy tulajdonságára két teljesen egyforma fájdalom épüljön rá. Vagyis két ugyanolyan derékfájdalmat érzek egyszerre. Ez nyilvánvaló képtelenség, hiszen semmi sem indokolja, hogy itt *két* fájdalomról beszéljünk. Az tehát teljességgel kizártnak tűnik, hogy pusztán numerikusan különböző mentális állapotok épüljenek rá ugyanarra a fizikaira, függetlenül attól, hogy hány személy van fizikailag valamilyen értelemben összekötve.

<sup>5</sup> Lásd pl. Chalmers 1996. 41–42, Ravenscroft 2005. 117–124, illetve magyarul Tőzsér 2009. 221–224.

A különleges hozzáférés elve tehát még a Farkas által vallott szofisztikált formájában is összeegyeztethetetlen azzal a nézettel, amely szerint a mentális logikailag ráépül a fizikaira. Ezzel persze semmi újat nem mondtam, aminek illusztrálására álljon itt az eddigieket kiegészítendő egy tekintélyérv is. David Chalmers a tudatosság kapcsán így fogalmaz: „A tudatosság megismerésében mutatkozó episztemikus aszimmetria világossá teszi, hogy a tudatosság kapcsán nem beszélhetünk logikai ráépülésről. Ha a logikai ráépülés lehetséges volna, nem volna ilyesfajta episztemikus aszimmetria; egy logikailag szuperveniciens tulajdonság esetében [...] semmilyen különleges szerepe nincs az első személyű megismerésnek” (Chalmers 1996. 102).

A két mentális állapot híve azonban esetleg egyfajta tulajdonságdualizmust védelmezhet, és azt mondhatja, hogy itt valójában arról van szó, hogy ugyanannak a fizikai állapotnak két különböző tulajdonsága az, amivel a két személy rendelkezik, mondjuk két külön fájdalomérzetet éreznek. Azonban itt is föltehetjük a kérdést: mi különböztetné meg e tulajdonságokat, miért beszélhetünk itt *két*, és nem egy tulajdonságról? Erre ellenfelünk esetleg válaszolhatná azt, hogy voltaképpen semmilyen kvalitatív különbség nincs a két tulajdonság között, ám valamiféle primitív partikuláris tulajdonságként, vagyis tovább elemezhetetlen trópusként értelmezve azokat, numerikusan mégis különböznek. Ám ekkor ugyanazt a kérdést tehetjük föl, mint az előbb: vajon a normális esetben is lehetséges, hogy én két numerikusan különböző, ámde minden tekintetben *ugyanolyan* fájdalomtulajdonsággal rendelkezzem egyszerre? Ez a feltevés számomra nemesak implauzibilisnek, hanem meglehetősen értelmetlennek is tűnik.<sup>6</sup>

Egyféleképpen persze nagyon is jól el tudjuk képzelni, hogy a két ember két különböző mentális állapottal rendelkezék, még hozzá úgy, hogy a szóban forgó fizikai állapot két különböző, ámde kvalitatíve megkülönböztethetetlen mentális állapotot *okoz*. Csakhogy a mentálisnak a fizikaitól való puszta oksági függése *inkompatibilis* a logikai ráépülés, és így a fizikalizmus tézisével. Az előző szakaszban bemutatott ellenpéldát igyekeztem úgy megkonstruálni, hogy ott a közösen birtokolt fizikai tárgy állapotán túl okságilag már semmilyen *fizikai* esemény vagy állapot fellépése ne legyen szükséges a mentális állapot bekövetkeztéhez. Ha tehát itt a több mentális állapot jelenlétének védelmében a mentálisnak a fizikaitól való oksági függésére hivatkozunk, akkor nem maradhatunk fizikalisták. Hiszen milyen különböző entitásokban okozna valamiféle változást a közös fizikai állapot megváltozása? Mivel az elképzelt példában minden releváns fizikai-fiziológiai tényező közös a két személynél, ezért a válasz csak az lehet, hogy a két személy valamiféle nem fizikai elemében – mondjuk az anyagtalan szellemi szubsztanciájukban – következik be ez a változás.

<sup>6</sup> E probléma részletes tárgyalásához lásd Újvári 2009. 116.

## KONKLÚZIÓ

A különleges privát hozzáférés mint a mentális megkülönböztető jegye tehát összeférhetetlen a fizikalizmussal. Voltaképpen ebben semmi újdonság nincs: legalább Frege óta ismerős az a gondolat, hogy a fizikai jelenségek szükségszerűen, lényegükénél fogva publikusak (Frege 1918/2000. 204–207). Farkas a mentális jelenségekkel kapcsolatban a privátság egy szofisztikált változatát vallja: bár maguk a mentális események és állapotok publikusak, az a mód, ahogyan hordozzuk hozzájuk fér, privát. Mármint ha helyes, e tanulmány érvelése azt mutatta meg, hogy a fizikai jelenségek lényegi publikusságába az is beletartozik, hogy nem lehet *ekvi* akadály a annak, hogy többen ugyanolyan módon ismerjék meg azokat.

Tanulmányomban tehát amellet érveltem, hogy a különleges hozzáférés Farkas által kidolgozott kritériuma maga után vonja a szubsztanciadualizmust. Régi közhely azonban, hogy ami az egyik filozófusnak *modus ponens*, az a másiknak *modus tollens*. Érvelésem tehát akár antifizikalista érvelés is lehet, ha az első opciót választjuk. Én azonban inkább a második lehetőség mellett dönténék: a szubsztanciadualizmust – itt ki nem fejthető indokok miatt – olyannyira elfogadhatatlan tézisnek tartom, hogy saját értelmezésem szerint érvelésem, ha helytálló, azt bizonyítja, hogy a különleges privilegizált hozzáférés nem lehet a mentális jelenségek megkülönböztető jegye.<sup>7</sup>

## IRODALOM

- Chalmers, David J. 1996. *The Conscious Mind*. Oxford, Oxford University Press.
- Farkas, Katalin 2008. *The Subject's Point of View*. Oxford, Oxford University Press.
- Frege, Gottlob 1918/2000. Logikai vizsgálódások. Első rész: A gondolat. Ford. Máté András. In uő: *Logikai vizsgálódások*. Budapest, Osiris. 191–217.
- Ravenscroft, Ian 2005. *Philosophy of Mind. A Beginner's Guide*. Oxford, Oxford University Press.
- Rorty, Richard 1970. Incorrigibility as the Mark of the Mental. *The Journal of Philosophy*. 67. 399–424.
- Tózsér János 2009. *Metafizika*. Budapest, Akadémiai Kiadó.
- Újvári Márta 2009. *Metafizikai dilemmák: Szubsztanciák, trópusok, változás*. Budapest, L'Harmattan.

<sup>7</sup> Ezúton szeretném köszönetemet kifejezni Tózsér Jánosnak, aki a tanulmány szövegének korábbi változatához fűzött hasznos megjegyzéseivel segített a végleges verzió megírásában.



## A privilegizált hozzáférés mint a mentális kritériuma: három aggály

*The Subject's Point of View* című könyvében Farkas Katalin azt a meglehetősen népszerűtlen karteziánus gondolatot védelmezi, hogy a mentális ismertetőjegye a privilegizált hozzáférés. A mentális állapotokat, tulajdonságokat vagy eseményeket az különbözteti meg azoktól az állapotoktól, tulajdonságoktól és eseményektől, amelyek nem mentálisak, hogy birtokosuk introspekción vagy reflexión révén megismerheti ezeket, míg más nem ismerheti meg ezeket ilyen módon. Bár ezt az álláspontot rendkívül világosan és felkészülten védelmezi, számomra nem sikerült csábítóvá tennie. Az alábbiakban három aggályt fogok röviden megfogalmazni. Nem látom, hogy (1) Farkas hogyan tudja mentálisnak tekinteni a diszpozicionális állapotokat, (2) hogyan adhat számot mások elméjének ismeretéről, és (3) hogyan kerülheti el a mentális állapotokkal kapcsolatos epifenomenalizmust.

(1) *Diszpozicionális állapotok*. Farkas tökéletesen tisztában van azzal, hogy számos olyan mentális állapot van, amelynek introspektíve nem vagyunk tudatában. A diszpozicionális állapotok, köztük a filozófusok első számú kedvencei, a hitek és a vágyak, fennállásuk időtartamának nagy részében nem tudatosak. Sőt, olyan hiteink és vágyaink is lehetnek, amelyeknek soha egy pillanatig sem vagyunk tudatában. Ha a mentális ismertetőjegye a privilegizált hozzáférés, akkor úgy tűnik, ezeket a nem tudatos állapotokat törölnünk kell a mentális állapotok listájáról. Ezt Farkas, nagyon helyesen, vonakodik megtenni. Egyesek olyan módon próbálják a nehézséget orvosolni, hogy azt állítják: ezek az állapotok is potenciálisan tudatosak, azaz képesek lehetünk reflexión útján felismerni őket. Valamilyen, számomra nem egészen érthető okból Farkas nem szimpatizál ezzel a megoldással, hanem a következőt feleli. A speciális hozzáférés, mint a mentalitás ismertetőjegye, a mentális *típusokat* azonosítja, a nem tudatos állapotok pedig nem képeznek külön típust. Az ismertetőjegye a hiteket, vágyakat és társaikat a mentális típusok közé sorolja, s ez tökéletesen összefér azzal, hogy e típusoknak vannak nem tudatos példányai is.

Ezt a választ nem érzem megnyugtatónak. Először is a szokásostól eltérő módon értelmezi „a mentális ismertetőjegye” (*mark of the mental*) kifejezést. Ez a

kifejezés Brentanótól származik – „az intencionalitás a mentális ismertetőjegye” –, s ezt Brentano példányokra, és nem típusokra vonatkoztatta, s ma is így szoktuk használni. De ha ezen túl is tesszük magunkat, akkor is van egy további probléma. Tegyük fel, hogy valaki a filozófusok ismertetőjegyének azt tekinti, hogy filozófia tanszéken tanítanak. De hát vannak filozófusok, akik nem tanítanak filozófia tanszéken. Ha az illető azt feleli, hogy az ismertetőjegy így is helyesen azonosítja a típust, ezt aligha fogadnánk el. A filozófus mint típus valóban azonosítható a filozófia tanszékekre való hivatkozással, a következő módon. A filozófus típusba azok tartoznak, akik bizonyos vonatkozásokban ugyanolyanok, mint azok, akik a filozófia tanszékeken tanítanak. De itt kilóg a lóláb: a filozófus ismertetőjegye valójában az, ami közös a filozófia tanszéken tanítóknak és azoknak, akik nem tanítanak filozófia tanszéken. A filozófia tanszéken tanítás tehát, bár gyakori és jellegzetes vonás, nem ismertetőjegy.

Erre Farkas felelhetné azt, hogy az analógia félrevezető, mert bizonyos esetekben minden további nélkül elfogadunk egy típus jellemzésére olyan vonásokat, amelyekkel a típus egyes példányai nem rendelkeznek. Az élőlénynek mint típusnak ismertetőjegye a szaporodás, jóllehet számos individuális élőlénynek nem sikerül szaporodnia.

Ez a válasz azonban nem elégséges a nehézség elhárításához, mégpedig a következő miatt. Bár vannak múltó hiteink és múltó vágyaink, hiteink és vágyaink közül sok tartós holmi. Azért érdemes igazolt hitekre szert tennünk, hogy ezeket szükség esetén bármikor segítségül hívhassuk cselekvéseink meghatározására, tehát ezeknek tartósan rendelkezésre kell állniuk. Hasonlóképpen, ha nem lennének tartós vágyaink, nem lehetnének hosszú távú terveink sem, hanem mindig csak pillanatnyi vágyainkat követhetnénk. Úgy vagyunk azonban megépítve, hogy egyszerre csak viszonylag kis számú dolognak lehetünk tudatában. Ennélfogva képtelenek vagyunk összes tartós hitünket és vágyunkat folyamatosan tudatunkban tartani. Hiteinknek és vágyainknak ezért nem olyan egyedi és esetleges balszerencséje, hogy többnyire nem vagyunk tudatunkban, mint amilyen egyedi és esetleges balszerencséje egy élőlénynek, ha nem szaporodik. Az, hogy hiteink és vágyaink jó részének nem vagyunk tudatában, kognitív berendezkedésünk természetéhez tartozik. Ha egy állítólagos ismertetőjegynek bizonyos dolgok nem alkalmilag és véletlenszerűen, hanem szisztematikusan és természetükből fakadóan nem felelnek meg, akkor az ismertetőjegy nem védhető meg pusztán azáltal, hogy azt mondjuk: az ismertetőjegynek a típusokat kell megragadnia, nem pedig minden egyes példányt.

(2) *Mások elméjének ismerete.* Első látásra nincs itt semmilyen probléma. Az, hogy a szubjektumnak privilegizált hozzáférése van saját mentális állapotaihoz, nem azt jelenti, hogy más nem férhet hozzá az ő mentális állapotaihoz, hanem csak azt, hogy más nem ismerheti meg őket ugyanolyan módon. Könyvének második részében Farkas azonban részletesebb elemzést nyújt a mentálisról, s ebben a mentálist a fenomenálissal, illetve a perspektívafüggővel azonosítja. Például:

[A] mentális tények perspektivikus tények, vagyis azonosságuk lényegi meghatározója, hogy ilyen és ilyen módon vannak a szubjektum *számára*. Az érzéki tapasztalatok fenomenális tulajdonságaival kapcsolatos tények – tudniillik amelyek azzal kapcsolatosak, hogy a dolgok milyenek látszanak, milyenek ízlenek és milyenek érződnek számunkra – paradigmaticus példái a perspektivikus tényeknek, mivel a látszat és az érzés mindig a szubjektum *számára* látszat és érzés, s ennél fogva egy bizonyos nézőpontot feltételez. [...] A tudatos mentális események fenomenális természetét privilegizált módon hozzáférhető számunkra. Az általam védelmezett internalista álláspont szerint ez kimeríti ezeknek az eseményeknek a mentális természetét (Farkas, Katalin 2008. *The Subject's Point of View*. Oxford, Oxford University Press, 130–131).

Úgy tűnhet, az, hogy a mentális fenomenális és perspektívafüggetlenség, még mindig nem okoz problémát, mert – mint Farkas nagyon világosan leszögezi – a mentális és a nem mentális nem zárja ki egymást. A határt a mentális és a *kizárólagosan* nem mentális között kívánja megvonni; például, míg a villámlás *kizárólagosan* nem mentális, fizikai esemény, addig a fogfájás fizikai esemény is és mentális esemény is.<sup>1</sup> Azért mentális, mert az illető, akinek fáj a foga, olyan módon fér hozzá, ahogy más nem: a fogfájás milyensége, fenomenális jellege csak az illető perspektívájából elérhető. Ebből nem következik, hogy én nem tudhatom, hogy az illetőnek fáj-e a foga. A fogfájáshoz mint fizikai eseményhez hozzáférék. Az viszont következik, hogy amikor tudom, hogy az illetőnek fáj a foga, nem tudom, hogy történik-e bármilyen *mentális* esemény.

Ez talán lerágott csontnak tűnik, de nem az. Vegyünk egy példát. X-et elhagyja a barátnője, aki aztán sűrűn cseréli a partnereit. X azt állítja, hogy mélyseges erkölcsi felháborodást érez a leányzó feslettsége miatt. Tegyük fel, hogy maximálisan őszinte, és szentül meg van győződve arról, hogy amit érez, erkölcsi felháborodás. Én azonban úgy gondolom, hogy X egyszerűen féltékeny. Farkas tisztában van azzal, hogy vannak elfojtott érzések, hiszen a könyv első részében gyönyörűen elemzi Anna Karenina önmaga elől is titkolt vonzalmát Vronszkij iránt. Most azonban az a kérdés, hogy a második részben kifejtett álláspont alapján, amely szerint az, ami mentális, fenomenális és perspektívafüggetlenség, mit is kezdhethünk ezekkel az esetekkel. Az ilyen esetekről azt szoktuk gondolni, hogy a két leírás, X-é és az enyém, összeegyeztethetetlen. Vagy erkölcsi felháborodást érez, vagy féltékenységet. De hogyan lehet itt ellentmondás? X érzelmi mint mentális állapotok perspektívafüggetlenség és fenomenálisak, s ilyen mivoltukban, azaz mentális állapotokként, számomra nem hozzáférhetők. Amikhez én hozzáférék, azok X érzelmi mint fizikai állapotok. Akkor miért ne volna lehetsé-

<sup>1</sup> A továbbiakban végig fizikalista módjára fogok beszélni. Erre Farkas felhatalmaz, hiszen hangsúlyozza, hogy álláspontja összeegyeztethető a fizikalizmussal.

ges, hogy az erkölcsi felháborodás mint mentális állapot azonos a féltékenységgel mint fizikai állapottal, amely esetben nincs ellentmondás?

(3) *Epifenomenalizmus*. A mentális állapotok internalista értelmezése mellett egyik standard érv a mentális okozással kapcsolatos. Azt szokás mondani, hogy a mentális állapotoknak azért kell internálisaknak lenniük, mert a mentális állapotok okozzák viselkedésünket. Egy állapot azonban csak belső tulajdonságai révén bír oksági erővel. Ha a mentális állapotok tartalma nem belső tulajdonság, akkor a mentális állapotok tartalma nem lehet okságilag hatékony. Farkas internalista, de nem használja fel ezt az érvet. Ez természetesen nem kifogásolható, csak hogy az az érzésem, hogy koncepciójába talán nem is illeszthető bele a mentális állapotok oksági hatékonysága. Vegyük megint X-et, s tegyük fel, hogy ő téved, és nekem van igazam: a volt barátnőjével kapcsolatos megnyilvánulásait nem erkölcsi felháborodása, hanem féltékenysége motiválja. Az én magyarázatom azonban az ő féltékenységére mint fizikai állapotra hivatkozik, hiszen az érzéseivel kapcsolatos mentális, azaz perspektívafügő tényekhez én nem férek hozzá. Ez általánosítható. Sem hétköznapi, sem tudományos magyarázataink során nem hivatkozhatunk perspektívafügő tényekre. Azt jelenti-e ez, hogy a mentális jelenségeknek mint mentális jelenségeknek, azaz perspektívafügő fenomenális vonásoknak nincs magyarázó potenciáljuk? Hogy a mentális események nem mentális mivoltuknál fogva, hanem fizikai mivoltuknál fogva magyaráznak, s a mentális mint mentális epifenomenális?

## Ahogy a dolgok tűnnek a szubjektum számára

Farkas Katalin szerint a „mentális” megkülönböztető jegye a privilegizált hozzáférés: azok az entitások mentálisak, amelyeket a szubjektum olyan módon ismerhet meg, amilyen módon azokat más nem ismerheti meg. Mi a privilegizált hozzáférés metafizikai alapja? Hogyan lehetséges, hogy van olyan része a valóságnak, amelyhez a szubjektum olyan módon fér hozzá, amilyen módon más nem tud? Farkas válasza:

Lehetséges magyarázattal szolgálhat az a megfigyelés, hogy a mentális tények *perspektivikus* tények. Szubjektumnak lenni nem más, mint *nézőponttal* rendelkezni. Az elmével rendelkező lényeket nemcsak körülveszik a dolgok, hanem *e dolgok meg is jelennek számukra bizonyos módon* (Farkas 2008a, 30–31. Kiemelések tőlem – T. J.).

Később Farkas így ír:

A tapasztalat mentális természete azáltal adott, *ahogyan a dolgok tűnnek számunkra*, amikor rendelkezünk ezzel a tapasztalattal. Az érzéki tapasztalat esetében ez a tapasztalat *fenomenális karaktere*, vagy az, *amilyen* a kérdéses tapasztalattal rendelkezni (Farkas 2008a, 88–89. Kiemelések tőlem – T. J.).

Farkas a következő álláspontot képviseli tehát: (1) valamennyi mentális állapot perspektivikus, értve ezen, hogy minden mentális állapotban a dolgok valahogyan tűnnek a szubjektum számára, és (2) az, ahogyan a dolgok tűnnek a szubjektum számára, nem más, mint a kérdéses mentális állapot fenomenális karaktere, vagyis amilyen abban a mentális állapotban lenni. Más szavakkal: Farkas azt, ahogyan a dolgok tűnnek a szubjektum számára, a szubjektum által instanciált fenomenális tulajdonságokkal *azonosítja*, és ennél fogva azt a kijelentést, hogy „minden mentális tény perspektivikus tény”, úgy érti, hogy minden mentális állapot *feltételez egy szubjektumot*, amely *átéli* a kérdéses mentális állapotot.

Az „ahogy a dolgok tűnnek a szubjektum számára” kifejezést azonban másképpen is lehet érteni. Például Alex Byrne szerint az, ahogyan a szubjektum számára a dolgok tűnnek, nem a kérdéses mentális állapot fenomenális tulajdonságaival, hanem a kérdéses mentális állapot intencionális/reprezentációs *tartalmával* azonos (Byrne 2001. 200–217). E javaslatnak is van plauzibilitása, hiszen a *dolgok* tűnnek valahogyan a szubjektum számára, és a módot, ahogyan a szubjektum számára a dolgok tűnnek, kézenfekvő a kérdéses mentális állapot tartalmának tekinteni.

Farkas és Byrne álláspontja a következőképpen viszonyul egymáshoz. Mindketten úgy gondolják, hogy egy mentális állapotot úgy kell leírunk, hogy a szubjektum számára a dolgok valahogyan tűnnek. Csakhogy míg Farkas e leírást úgy olvassa, hogy „a *szubjektum számára* valahogyan tűnnek a dolgok”, addig Byrne úgy, hogy „a *dolgok valamilyennek tűnnek* a szubjektum számára”. Míg Farkas a dolgok valaki számára való valamilyen módon történő megjelenését vagy tűnését azzal azonosítja, amilyen átélni a szubjektum számára ezt a tapasztalatot, addig Byrne úgy tekinti, hogy a dolgok valaki számára való valamilyen módon való megjelenése vagy tűnése irányul valamire, van intencionális/reprezentációs tartalma, és a világot reprezentálja valamilyen módon.

E nézetkülönbségnek komoly következményei vannak. Farkas a könyvében (és máshol, például Farkas 2008b) következőképpen gondolkodik: a szubjektum által instanciált fenomenális tulajdonságokon van a hangsúly, ennél fogva ezek határozzák meg a mentális állapotok intencionalitását. (Legalábbis ha rendelkeznek ilyesmivel. Farkas szerint nem minden mentális állapot intencionális, lásd Farkas 2009.) Byrne ezzel szemben úgy okoskodik: a szubjektum mentális állapotainak az intencionális/reprezentációs tartalma az elsődleges, ennél fogva ezek határozzák meg mentális állapotaink fenomenális karakterét, ezeken szuperveniálnak a fenomenális tulajdonságok. Így fogalmaz: „nem lehet különbség a fenomenális karakterben anélkül, hogy ne volna a tartalomban” (Byrne 2001. 204).

Álljunk meg azonban egy pillanatra. Nem azt akarom mondani, hogy Byrne álláspontja plauzibilisebb Farkasénál. Ha történetesen Byrne tanulmányáról írnék, előszeretettel hivatkoznék Farkas álláspontjára mint plauzibilis alternatívára. Így jár el Tim Crane is, amikor Byrne-t bírálja:

Nos, bizonyosan plauzibilis azt mondani, hogy „a mód, ahogyan a dolgok tűnnek a szubjektum számára”, olvasható úgy, mint ami a tapasztalat tartalmával szinonim. De valaki [...] mondhatja, hogy „a mód, ahogyan a dolgok tűnnek a szubjektum számára”, más módon is olvasható: nevezetesen, mint ami *magának* a tapasztalatnak a fenomenális karakterére vonatkozik (Crane 2008. 287, kiemelés az eredetiben).

Nem azt akarom tehát állítani, hogy az egyik álláspont plauzibilisebb a másiknál, hanem azt, hogy mindkettőnél van sokkal plauzibilisebb. Nevezetesen: az „ahogyan a dolgok tűnnek a szubjektum számára” kifejezésben *mind* a dolgok (*amelyek* megjelennek), *mind* pedig a szubjektum (amely *sámára* a dolgok megjelennek) egyformán hangsúlyos.

A következőt állítom: minden megjelenés *valaminek* a megjelenése (ennyiben minden megjelenés rendelkezik intencionális tartalommal), és minden megjelenés *valaki számára* való megjelenés (ennyiben minden megjelenés rendelkezik fenomenális karakterrel). Más szavakkal: minden megjelenésnek van világra irányuló aspektusa *is* (azaz a világot jeleníti meg, a *világot reprezentálja* valamilyen módon), de éppúgy van fenomenális aspektusa *is* (azaz a világot a *szubjektum számára* jeleníti meg, azaz valamilyen a szubjektum számára). Colin McGinn szavaival:

A [...] tapasztalatok Janus-arcúak: a külvilágra irányulnak, de szubjektív arcukat mutatják a szubjektum számára, és valamilyenek a szubjektum számára. De e két arc kifejezése nem tér el egymástól: a tapasztalat valamilyensége annak a függvénye, amire a tapasztalat irányul, és amire a tapasztalat irányul, az valamilyenségének a függvénye (McGinn 1991. 29–30).

E perspektívából nézve mind Farkas, mind Byrne álláspontja félmegoldás: mindkettő megragadja a mentális állapotok egyik alapvető karakterét, miközben a másiktól megfosztja azt.

## IRODALOM

- Byrne, Alex 2001. Intentionalism Defended. *Philosophical Review*. 110. 199–239.
- Crane, Tim 2008. Intentionalism. In Ansgar Beckermann – Brian McLaughlin (szerk.) *Oxford Handbook of Philosophy of Mind*. Oxford, Oxford University Press. 474–494.
- Farkas, Katalin 2008a. *The Subject's Point of View*. Oxford, Oxford University Press.
- Farkas, Katalin 2008b. Phenomenal Intentionalism without Compromise. *Monist* 91. 273–293.
- Farkas, Katalin 2009. Not Every Feeling is Intentional. *European Journal of Analytic Philosophy* („Challenges for Representational Theories of Mind and Consciousness”, Monographic issue edited by Tamás Demeter and János Tőzsér). 5/2. 39–52.
- McGinn, Colin 1991. *The Problem of Consciousness*. Oxford, Blackwell.

## Válasz Márton Miklósnak, Forrai Gábornak és Tózsér Jánosnak

(1) Könyvem egyik fő tétele, hogy a mentális jelenségek megkülönböztető jegye a privilegizált hozzáférés, és Márton Miklós pontosan rekonstruálja, hogy mit értek ezalatt. A sajátos hozzáférés elvéből, amennyire látom, önmagában nem következik a dualizmus, vagyis az az állítás, hogy az elme nem fizikai létező. Márton Miklós szerint viszont következik. Az az igazság, hogy ebben a kérdésben a véleményemet korántsem mondanám véglegesen rögzítettnek; egyelőre úgy látom, hogy a fizikalizmus összefér a sajátos hozzáférés tételével, de nem omlana össze a világ, ha az ellenkezője derülne ki. Ez az esemény azonban még nem következett be, ugyanis Márton Miklós érve, bár igen érdekes, egyelőre nem győzött meg.

Márton gondolatmenete a következő. Tegyük fel, hogy elfogadjuk a fizikalizmus tételét: a mentális események fizikai eseményekkel azonosak. Egy fájdalomélmény például azonos valamilyen agyi eseménnyel. Képzeljük el a következő helyzetet: két embernek megsérül az az agyterülete, amely bizonyos típusú fájdalomélményekért felelős, a jövő fejlett idegsebészete pedig lehetővé teszi, hogy egy implantátummal helyreállítsák a sérült rész működését. Az implantátum, megfelelően összekötve az agy többi részével, átveszi a sérült terület funkcióját. Képzeljük el továbbá azt, hogy egy bizonyos esetben két embernél használnak egyetlen implantátumot, ami mindkettejük agyához megfelelően kapcsolódik. (Talán ahhoz hasonlóan képzelhetjük el ezt a felállást, mint amikor összenőtt ikrek megosztanak egy szervet.) Mivel a fizikalizmus szerint a mentális események azonosak a fizikai eseményekkel, egyetlen fizikai esemény lesz azonos mindkét alany mentális eseményével. És ha van privilegizált hozzáférés, akkor mindketten ugyanahhoz az eseményhez férnek hozzá privilegizált módon: egy olyan eseményhez, amely nem csak a saját, hanem egyben egy másik ember mentális eseménye. Vagyis nem áll, hogy mindenkinek csak a saját mentális eseményeihez van sajátos hozzáférése.

Szerintem nem stimmel az érv. Elképzelésem szerint a reflexió egy sajátos megismerési képesség, amely bizonyos állapotokról vagy eseményekről tudást eredményez. A tudás propozicionális. A reflexió eredménye annak tudása, hogy az alany rendelkezik valamilyen mentális vonással. A tudás tárgya lehet például,



hogy fáj valamim, vagy hogy homályosan látok, vagy hogy most éppen Descartes-ra gondolok. Az első személyű tudás lényeges eleme, hogy az alany magának tulajdonít vonásokat: nemcsak arról szerzek tudomást, hogy valakinek fáj valamije, hanem hogy *nekem* fáj valamim. Az előbbi sem érdektelen információ, de az utóbbi az, aminek számomra igazán jelentősége van. Amikor azt állítom, hogy mindenki a saját mentális élményeihez fér hozzá sajátos módon, akkor ezalatt azt is értem, hogy *mint* sajátjához fér hozzá.

Az implantátumon osztozkodó két alany, nevezzük őket A-nak és B-nek, különböző tudásra tesz szert reflexió révén: A azt tudja meg, hogy neki, vagyis A-nak fáj valami, míg B azt, hogy neki, vagyis B-nek fáj valami. Egyiküknek sem lesz privilegizált hozzáférése ahhoz a tényhez, hogy a másiknak fáj valamije. Ezt egyébként Márton is megjegyzi, de szerinte „e különbség ismeretelméleti és nem metafizikai különbség: nem érinti azt a tényt, hogy ugyanazt a jelenséget ismerik meg ugyanolyan módon”. Nem tudom pontosan, hogy hogyan kell felosztani a területet itt az ismeretelmélet és a metafizika között, hiszen a privilegizált hozzáférés maga ismeretelméleti kérdés. Mindenesetre a két reflexió tudás tárgya két metafizikailag is különböző tény: az, hogy A-nak fáj valami, és az, hogy B-nek fáj valami.

A következő gondolat spekulatív, de számomra valószínűnek tűnik, hogy a fájdalomérzést megvalósító fizikai esemény nem korlátozódhat pusztán az implantátumban végbemenő folyamatokra. Egy mentális esemény egy szubjektum tudatfolyamában megy végbe; nincsenek önálló fájdalomeseemények szubjektum nélkül. Felteszem, hogy ez tükröződik a fizikai megvalósulásban is: az implantátum akkor valósít meg egy mentális eseményt, ha megfelelően be van „kötve” egy funkcionáló agyba. Azt gondolom, hogy legalább a kötések is hozzátartoznak a mentális jelenséget megvalósító fizikai eseményhez; még valószínűbb, hogy az agy nem annyira moduláris, hogy ez az esemény teljesen független legyen az agy többi részében zajló eseményektől. Ez azt jelenti, hogy a két mentális eseményhez két fizikai esemény tartozik.

(2) Forrai Gábor észrevételei közül először a második, a mások elméjének ismeretével kapcsolatos felvetést tárgyalom. Miként azt Forrai is megjegyzi, nézeitem szerint lehetséges mások mentális állapotairól tudást szerezni; a privilegizált hozzáférés nem azt jelenti, hogy csak én férhetek hozzá mentális állapotaimhoz, hanem az, hogy csak én férhetek hozzájuk egy bizonyos *sajátos* módon, vagyis reflexió segítségével. Mások is hozzáférhetnek, csak nem reflexióval. Forrai ezt úgy értelmezi, hogy szerintem mások mentális eseményeihez *mint* mentális eseményekhez nem férünk hozzá, hanem csak *mint* fizikai eseményhez (itt Forrai is támaszkodik arra, hogy megengedő vagyok a fizikalizmussal kapcsolatban). Vagyis csak azt tudhatjuk egy másik emberről, hogy egy bizonyos fizikai esemény megy végbe benne; ez történetesen azonos egy mentális eseménnyel, de mi ezt a körülményt nem ismerhetjük.

Én őszintén szólva egyáltalán nem így képzeltem el mások elméjének ismeretét. Amikor azt mondom, hogy mások mentális állapotairól tudást szerezhettünk, akkor ezalatt pontosan azt értem, hogy az olyanfajta harmadik személyű állítások, mint „Anna kétségbe van esve” vagy „Annának fáj valamije”, tudás tárgyai lehetnek. Csakhogy ennek a tudásnak a forrása nem a reflexió, mivel más emberek elméjére nem tudunk reflektálni. Mások elméjéről alkotott tudásunk egyik leggyakoribb forrása az a beszámoló, amit tőlük hallunk. Például valaki elmeséli, hogy egész nap fel volt dobva, mert nagyon megörült egy jó hírnél reggel, és már alig várja, hogy elmondhassa a legjobb barátjának – ebből én mindenféle dolgot megtudok az ő mentális állapotairól. Szó sincs arról, hogy én ezekről az állapotokról *mint* fizikai állapotokról szereztem volna tudomást. Ha van is olyan fizikai állapot, ami ténylegesen azonos a *feldobottsággal*, meg az *alig várással* (mondjuk valami agyállapot), azokat én biztosan nem tudom azonosítani, azt viszont nagyon is jól tudom, hogy az illető milyen mentális állapotban van. De világos, hogy az ő tudása meg az én tudásom ezekről a mentális állapotokról különböző forrásból fakad: az övé a reflexióból, az enyém a beszámolóból.

Én ezzel a felfogással kapcsolatban semmi problémát nem látok. Miért ne ismerhetné meg két ember ugyanazt a tényt két különböző megismerőképesség segítségével? Talán akkor jelenthetne nehézséget ez az elképzelés, ha a két alany közül az egyik teljes mértékben nélkülöznél az egyik fakultást és a vele járó tapasztalatokat. Például érvelhetnénk amellelt, hogy a látás révén szerzett tudást nem tudjuk teljes egészében közvetíteni szavak segítségével egy olyan embernek, akinek soha nem voltak bizonyos fajta látási tapasztalatai. De feltevésem szerint ez a helyzet nem hasonlít arra, amikor más emberek mentális állapotait próbáljuk megismerni. Mivel én is voltam már feldobva, meg örültem már jó hírnél, pontosan értem, hogy mit is mond a másik, amikor a saját mentális állapotairól így számol be.

Persze felvethetné valaki, hogy mégsem lehetünk teljesen bizonyosak abban, hogy a másik embernek olyasfajta fenomenális élményei vannak, mint nekem. Az is lehet, hogy egy beszámoló révén soha nem lesz olyan, minden részletre kiterjedő tudásom a másik ember elméjéről, mint amilyen neki van. Ez valóban így van, abszolút bizonyosságot és teljes megismerést itt nem érhetünk el, de ez nekem nem is céloim. Bár Descartes elmefelfogását nagyon nagyra értékelem, a karteziánus ismeretelmélet távol áll tőlem. Jó okom van azt hinni, hogy embertársaim érzéseit fiziológiai állapotaik határozzák meg; mivel hasonlóan konstruált lények vagyunk, jó okom van azt is hinni, hogy hasonló körülmények hasonló érzéseket váltanak ki belőlünk. Általában kevésbé győz meg a szkepticizmus, és a más elmékkel kapcsolatos szkepticizmus sem gyakorol rám nagyobb hatást, mint a többi változat.

Éppen ezért nem osztom Forrai Gábor azon aggályait, amelyek az említett feltevéseken alapulnak: mások mentális állapotait csak mint fizikai állapotokat ismerem meg. Forrai azonban egy másféle problémát is megfogalmaz, a nem

tudatos állapotokkal kapcsolatban. Kétségtelen, hogy a nem tudatos vagy tudattalan mentális állapotok léte első látásra okoz bizonyos problémákat a sajátos hozzáférés tézise számára. Valószínűleg ez a kérdés még további átgondolást is igényel, de azért vázoló az alapvető elképzelést. Miként a könyvben is jelzem, én rokonszenvezem azzal a nézettel, hogy a hitek és a vágyak soha nem tudatosak, mert ezek diszpozíciók, amelyeknek azután vannak megnyilvánulásai a tudatfolyamban. Például annak az állandó *hitemnek*, hogy a náthára nem használ a C-vitamin, a mostani azonos tartalmú tudatos *ítéletem* – ahogy a szavakat írom vagy mondok – az egyik megnyilvánulása; de a hit éppúgy nem azonos a tudatos ítélettel, mint ahogyan a diszpozíciót más esetekben is megkülönböztetjük manifesztációjától.

Az állításom azonban az volt, hogy annak ellenére, hogy a hit nem része a tudatfolyamnak, rendszerint mégsem okoz problémát számomra, hogy látszólag közvetlenül reflektáljak hiteimre. Ha valaki megkérdezi, hogy hiszem-e vagy sem, hogy a náthára használ a C-vitamin, erre rögtön tudok válaszolni, még hozzá úgy látszik, hogy valamiféle introspekció alapján; ugyanis ha azt kérdezik tőlem, hogy Forrai Gábor vagy bárki más hiszi-e ugyanezt, akkor erre a választ egész másképpen fogom keresni. Vagyis itt is megmutatkozik az első és a harmadik személyű hozzáférés különbsége. Nem tagadom, hogy teljes mértékben adós vagyok egy olyan elmélettel, ami ezt a jelenséget magyarázza – a könyvben nem vállalkozom az introspekció elméletének kidolgozására. Mindössze az aszimmetrikus hozzáférés jelenségére hívom fel a figyelmet.

Forrai úgy érti, hogy a diszpozicionális állapotokkal kapcsolatos lehetséges problémára azzal válaszolok, hogy a sajátos hozzáférés mentális típusokat azonosít, és ez „összefér azzal, hogy e típusoknak vannak nem tudatos példányai is”. A helyzet az, hogy én a típus/példány manővert nem ebben a kontextusban alkalmazom. A dolgozatban megkülönböztetem a nem tudatos (*non-conscious*) és a tudattalan (*unconscious*) állapotokat, és a manővert az utóbbiaknál alkalmazom, az előbbieknél nem. Ez csak terminológia, és talán nem is a legjobb, de a megkülönböztetés tartalma szerintem megállja a helyét. Ahogyan korábban mondtam, a diszpozicionális állapotok, mint a hitek és vágyak, nem tudatosak, a legtöbb esetben mégsem jelent problémát az, hogy közvetlenül, sajátos első személyű hozzáféréssel azonosítsam őket. Viszont vannak problematikus esetek: ilyenek például az önbeesapás esetei, valamint – ha vannak ilyenek – a freudi elnyomott vágyak, és ezeket nevezhetjük talán „tudattalannak”. Ezek remélhetőleg a nem tudatos állapotoknak csak egy (nem túl nagy) részhalmazát képezik. Így egyetértek Forraival abban, hogy hiteinknek és vágyainknak nem csak „egyedi és esetleges balszerencséje”, hogy nem tudatosak; ellenkezőleg, „az, hogy hiteink és vágyaink jó részének nem vagyunk tudatában, kognitív be rendezkedésünkhöz tartozik”. Ám úgy gondolom, hogy az egyedi és esetleges balszerencse, ha egy adott vágyhoz vagy hithez nem férünk hozzá az önbeesapás vagy a represszió miatt.

(3) Tőzsér János javaslata szerint valahányszor megjelenik valami egy szubjektum számára, a szóban forgó mentális esemény kettős természetű: egyrészt van intencionális tartalma, jelesül az, ami megjelenik, és van fenomenális karaktere, jelesül az, ahogyan a szubjektum számára a tartalom megjelenik. Tőzsér szerint az én álláspontom túlságosan egyoldalú, mert a kettős természetnek csak az egyik oldalát, a fenomenális karaktert hangsúlyozza.

Ahogy Tőzsér is jelzi, az egész kérdés abban az összefüggésben merül fel, hogy mi a mentális ismertetőjegye. Lehet persze azt is mondani, hogy a mentálisnak nincs ismertetőjegye, hanem a mentális jelenségek egy közös lényeggel nem rendelkező vegyes csoportot alkotnak. Ezt a kérdést most nem tudom érdemben tárgyalni, úgyhogy csak a gondolatmenet kedvéért tegyük fel, hogy igenis van ilyen ismertetőjegye. Ha minden mentális állapotnak van intencionális tárgya, ami valahogy megjelenik az alany számára, akkor talán tényleg érdemes elgondolkozni, hogy az elme természetében melyik az alapvetőbb, és talán érdemes azt elfogadni, hogy mindkettő egyformán alapvető.

Csakhogy, mint azt Tőzsér is megjegyzi, szerintem nincsen minden mentális eseménynek intencionális tárgya. Erről a könyvben nincs igazán szó, de azért igyekeztem úgy megfogalmazni az elméletet, hogy ez a lehetőség nyitva maradjon (és egy Tőzsér által hivatkozott másik tanulmányban érvelni is igyekszem az álláspont mellett). Tőzsér idézi Colin McGinn azon megállapítását, hogy a tapasztalatok kettős természetűek: a kívülágra irányulnak, és egyszersmind valamilyenek az alany számára. De szerintem ez nem minden tapasztalatra igaz, mert vannak olyan tapasztalatok, amelyek nem irányulnak a kívülágra vagy bármi másra, hanem csak egyszerűen *érzések*. Ilyen lehet például egy bizonyos fajta szédülés, egy tompa fáradtság, vagy egy általános, irányultság nélküli eufória. Ezekben az esetekben a tapasztalat egész jellegét az érzés adja meg, az, hogy milyen is az alanynak abban az állapotban lenni; a fenomenális jelleg nem valamilyen tárgy megjelenése, hanem pusztán az alany tudatának egy sajátos módosulása. Nem merül fel a kérdés, hogy mi az alapvetőbb, a fenomenális jelleg vagy az intencionális tárgy, mivel csak az előbbi van, az utóbbi nincs.

Az én gondolatmenetem, amit Tőzsér is idéz, a következő: a mentális ismertetőjegye a privilegizált hozzáférés; ennek metafizikai alapja a mentális tények perspektivikus természete; ez a körülmény pedig azonos azzal, hogy a mentális állapotoknak fenomenális jellege van. Ha jól értem, Tőzsér a harmadik lépés ellen emel kifogást: szerinte a perspektivikus jelleget a fenomenális karakterrel és intencionális irányultsággal kellene magyarázni. Csakhogy szerintem ez nem lehetséges, ugyanis az intencionális irányultság bizonyos esetekben hiányzik; ezzel szemben a fenomenális jelleg mindig jelen van. Ha tehát egy teljesen általános magyarázatot akarunk adni a privilegizált hozzáférés alapját jelentő perspektivikus jellegre, akkor ezt csak a fenomenális jellegben kereshetjük.

BOROS GÁBOR

GÁBOR GYÖRGY – VAJDA MIHÁLY (SZERK.)

## A lét hangoltsága.

*Tanulmányok a tudás sokféleségéről.* Budapest, Typotex.

2010. 369 oldal (Saeculum Könyvek)

*Dancing on the Edge*, kicsit szabadon: „borítóvalén táncoló” a címe a borítótervet ihlető képzőművészeti alkotásnak: egy sűrű (albatrosz?) kötelet húz magával repülés közben, s e kifeszült kötélén buddhista szerzetes (?), kung-fu harcos (?) kötélutános egy keréken egyensúlyoz. Nem volt sokkal könnyebb a szerkesztők feladata a kötet összeállításakor, hiszen a maguk elé kitűzött cél a korunkban lehető legátfogóbb filozófiai célkitűzés, s e célt megvalósítani nem saját kútfejükre építve vagy a tetszésük szerint válogatott szerzőkkel kellett, hanem egy meghatározott filozófiai műhelynek, az MTA Filozófiai Kutatóintézetének 2009-ben volt munkatársaival és vendégelőadóival. A cél a lét hangoltságának megragadása, nyilvánvaló allúzióként a *Sein* és a *Stimmung* heideggeri fogalmaira, amiből következik, hogy valamiképp a hajdani *metaphysica generalis* és *specialis* örökségképp a lét nagy tartományainak illeszkedéséről, összehangoltságuk módozatairól kell beszélni. Az ezt megközelítő tudás sokfélesége ebben a hagyományban magától értetődik, az arisztotelészi érzelmfilozófia ontológiai „habilitációja” Heidegger *Lét és időjében* (lásd Vajda Mihály kötet-címadó írását; azért nem rehabilitáció, mert az érzelmek korábban nem „lakoztak” az ontológiában), a megértés fogalmának létértelművé tágítása, a tudás laboratóriumi elszigeteltséghez kötődő koncepciójának

elvetése épp ezt jelenti. A tudás egyszersmind konstitutív is: a szöveggént, a leírt valójában megjelenő tudás nem önmagukban létező valóságelemekhez kapcsolódik, hanem maga tesz kísérletet a faktumként adott lét dimenzionálására, tördelésére, hogy egy-egy sajátos perspektíva felől mutakozzék meg a lét hangoltsága. E perspektivitáshoz az is hozzátartozik, hogy az adott lét hangoltsága bizonyos nézőpontokból *jól*-hangoltságként („*wohltemperiert*”), más nézőpontokból bizony inkább *el*hangoltságként „hangzik”.

A recenzió maga természetesen a legperspektivikusabb műfaj, az adott kötet hangoltságát a recenzens saját nézőpontjából vizsgálja, amely nem azonos sem a szerzők, sem a szerkesztők, sem az olvasók nézőpontjával. A recenzens, előrebocsátva véleményét, úgy „hallja”, a szóban forgó kötet összességében jól-hangolt, noha *ez* nyilvánvalóan nem jelenti, hogy minden frással, minden állítással egyetért.

Talán meglepő, de most úgy látom, a kötet kulcsa Sziklai László „*Minden papírdarabkára vadászni...*” című írása, amellyel az a tudás, amelynek sokféleségéről szó van, önreflexívvé válik, szituálódik. Ebben az írásban ugyanis a Filozófiai Kutatóintézet létrejöttéről, első időszakáról van szó, inkább nem akarva, mint akarva, kitekin-téssel a jövőre, ami éppenséggel a kötet jelenideje. Ennek az írásnak a nyomán úgy

tűnik, a Lukács Archívum és a Filozófiai Kutatóintézet kétszeresen is kötél-táncosként lebeg(ett) az ő semmije, vagyis nemléte fölött. Az egyik a Sziklaitól is exponált probléma, hogy vajon mit jelent archívumot létesíteni az önmagát nagy, szubsztantív világtörténelmi áramlatként megragadó marxizmus-leninizmus egy vezető gondolkodója számára, akinek – egy lehetséges értelmezés szerint – életfeladata abban is állhatna, hogy a maga személyes individuum voltát visszazorítva az „osztály” világtörténelmi individualitásában oldódják fel. (Ebben a problémában csíráként benne rejlik a kötetnek csaknem egészét látenszen strukturáló probléma a történelem kétféle megközelítéséről: vajon a történelem annak feltárása-e, *wie es gewesen*, s akkor ehhez a filozófiától módszertani fogódzók lehet remélni; vagy abból indulunk ki, hogy van tudásunk az emberi történelem emberen túli nagy sodrásáról, akár van ennek emberen túlian személyes irányítója, akár nincs, akár van kivehető, esetleg valamiképp előre jelzett iránya, akár nincs. Ha ez utóbbi állásponton vagyunk, akkor nem módszertani fogódzók kell keresnünk a filozófiában, hanem megerősítéseket a jól dokumentált tényektől s feltárásuk módjától függetlenül tudott vagy legalább sejtett fő áramlatról.) Ha feloldódik, akkor a tényeket rögzíteni próbáló cetlivadászatnak aligha van értelme, ha nem oldódik fel, akkor viszont túl önállóvá válik, önmaga törvényét követi a világtörténelem állítólagos törvényei helyett, melynek az adott (volt?) párt a nagy tudója, s ez esetben elveszti az igényjogosultságát arra, hogy épp e mozgalom vezető gondolkodójának számítsa. Vagyis egy világosan artikulálható alternatíva bármely ágát követve is oda jutunk, hogy nincs létjogosultsága az archívumnak. A valóság, a megoldás a lavírozás lett a pártakat és az utóbbi egy-két évszázad európai kultúrájának nagy individuumok iránti tisztele-

te között, amely a személyes archívumok intézményét életben tartja. Mármint így rekonstruálható ez a mából visszatekintve, akkoriban nyilván sokkal pragmatikusabban s gyakorlatibb szinten zajlottak az események, amelyekre nagyon jól rávilágítanak a Sziklai-tanulmány jegyzeteiben közölt dokumentumok: például a Radnóti Sándor kibetűzői alkalmazásának meghosszabbításáért folyamodó levél. E gyakorlati értelemben vett lavírozás mellett meg kell természetesen említeni ennek elméleti oldalát, amelyet a reformmarxizmus névvel szokás illetni. S ez a fent említett kétszeres kötél-táncosi helyzet másik alproblémája. Tudniillik hogyan lehetne, s miért kellene Filozófiai Intézetet alapítani és fenntartani egy olyan országban, amelyben a vezető párt hivatalos ideológiai öntudata szerint olyan mozgalmat szolgál s irányít, amely meghaladta a klasszikus értelemben vett filozófiát, s amelynek ráadásul minden jel szerint (bár talán valamennyire maguk ezek a jelek is hozzátartoznak ehhez vagy ahhoz az ideologikus öntudathoz, lásd Simon József: *Filozófia vagy irodalom?* In Boros Gábor [szerk.] *A hetvenes évek filozófiai lehetőségei és valósága Magyarországon és az NDK-ban*. Budapest, L'Harmattan, 2010, 216–222) a hagyományos, kommunizmus–szocializmus–marxizmus uralma előtti korszakokban sem volt akkora jelentősége a kultúrában, hogy egy ilyen önálló kutatási intézményt indokolttá tett volna? A „reform” mint minősítő szócska a „marxizmus” mellett ebben az összefüggésben kétségkívül megkérdőjelezést jelentett: *filozófiai* intézetet alapítani és fenntartani közel állt ahhoz, amit a klasszikus európai filozófiatörténet önálló művelése jelentett nagyjából ugyanebben az időszakban: a marxizmus-leninizmus ideológiájának legalább látens megkérdőjelezését.

Nem szeretnék hosszabban időzni ennél az írásnál. Azért emeltem ki kulcsként, mert a nemlétezés – Lenin Intézet – Filo-

zófiai Intézet és Lukács Archívum vonalat meg lehet hosszabbítani, s annak (remélhetőleg pusztán az idő természetéből fakadóan pillanatnyi) végpontján az a Filozófiai Kutatóintézet áll, amelynek 2009. évi kutatói terméséből állt össze a jelen kötet. Im már nem élehetetlen, komolyan nem vehető szélsőségek közti lavírozásra épülve – akár gyakorlati, akár inkább elméleti értelemben vesszük ezt –, hanem az európai kultúra elengedhetetlen alkotóelemét jelentő, hagyományos – azaz hagyományától, történelmétől elválaszthatatlannak tekintett – filozófia művelésének helye. Lukács György marxizmusa csakúgy, mint korai „szellemtörténeti” kötődése egyáltalán nem valamiféle megoldhatatlannak tetsző paradoxon már, hanem az a feladat, hogy mindezek a mozzanatok beépíteszenek a filozófia hagyományáról alkotott képünkbe, illetve hogy saját filozófiai próbálkozásainkhoz alapul szolgáljanak. Ezt a feladatot a Lukács Archívum munkatársai értelemszerűen inkább Lukács oldaláról, kissé a *wie es gewesen* oldaláról közelítik meg (bár nem kizárólag, példa erre Hévízi Ottó írása), míg a szorosan vett Kutatóintézet munkatársai a filozófiai hagyomány és a saját filozófiai törekvések oldaláról (ez sem kizárólagos, példa rá Gángó Gábor írása).

Mint említettem, a kötet látens, általános strukturáló elve a történelem kétféle megközelítésmódja. Elég egyértelműen jelenik meg e kettősség az első nagyobb egységben, melynek címe is utal a történelemre: „A történelemhez tartozunk, mielőtt történelmet íránk” (természetesen mást jelent a „történelem” az első előfordulásban, mint a másodikban). Gábor György *Az ítrék – Róma és Jeruzsálem*, Frenyó Zoltán *Hittita és kényszer Szent Agoston antidonatista leveleiben* és Turgyoni Zoltán *Vannak-e európai értékek?* című írása képviseli a szubsztantív értelmű történelem-felfogást, míg a történész vendégelőadók írásai – Gyáni Gábor *A történelmi tapasztalat fogalmának historizálása* és Sonkoly

Gábor *Kulturális örökség és történelem a cárosban* című írása – a történelemnek mint olyan tapasztalati térnek a létét posztulálják, amely mögé, alá vagy fölé műndszeriani vagy más okokból nem juthatunk. Az értékek megjelenése világ(üdv)történelmi perspektívából ez utóbbi írásokban nehezen képzelhető el, míg az elsőként említettekben elengedhetetlen. Mester Béla írása *A felső hatalmasságok fogalma a 16. századi magyar reformáció politikai gondolkodásában* címmel, történészi alapossággal épp a politikum létében megtapasztalható történelmi ténynek a történelem fölötti, személyes isteni szférába való teoretikus bekapcsolásának jár utána a reformáció körüli vitákban, különös tekintettel a magyarországi reformáció különböző képviselőire. Ezzel üdvös módon kitágul a magyarországi filozófiai gondolkodás fogalma időben és szellemi térben is, nagyjából Quentin Skinner *Foundations of Modern Political Thought* című alpmunkájának folytatásaképp. S ha már a kora újkornál tartunk, a recenzens nagy örömeire még két izgalmas tanulmány olvashatunk e korról, Szántó Veronika *Harrington és a politikai test vérkeringése* és Székely László *Az orchideák bolygója: Pascal végtelen térségei és a Földön kívüli élet* című írását (mindkettőt az „Egy felfoghatatlan mindenségbe való belevetetés” című részben találjuk). Szántó írása is kapcsolható az általános strukturáló elvhez: a 17. századi vita a vér vagy a szív központi jelentőségéről az emberi testen belül teljességgel politikai színezetet kap: a szív a hatalomkoncentráció jelképe, *azé az* egy, kitüntetett ponté, amelyen keresztül az, ami nem e világról való, mintegy „felső hatalmasságként” hatást gyakorolhat e világ történéseire. Az éppen „felfedezett” vérkeringés mint az egész testműködést összefogó funkció révén viszont a vér szert tett egy alternatív jelképi szerepre, a hatalomkoncentrációval szemben a hatalomdiszt-

ribúcióéra – noha azt azért mégsem lehet állítani, hogy egyértelmű megfelelés lett volna a szívet hangsúlyozó medikusok és királyságiak közt egyfelől, és a vért hangsúlyozók és republikánusok közt másfelől. Székely László érdekes módon nem a történelem síkjának bejárását-megtapasztalását folytatja, hanem a világegyetem tereit „járja be”: az Istentől elhagyottnak tekintett pascali világegyetemben más értelmes lényeket kereső újkori filozófusokat és 20. századi filozófáló természettudósokat és irodalmárokat vizsgál.

De annyi bizonyos, hogy a vér kiemelését a korban úgy fogták fel, mint ami megnehezíti, ha lehetetlenné azért nem is teszi a transzcendens kötődésű, halhatatlan individuális lélek hozzákapcsolását a testhez. Ez a „mortalizmus” bukkan fel Mesterházi Miklós Kant-értelmezésében – *A metafizika álmai, megmagyarázva egy szellemlátó álmai által* –, aki stilsztikailag is izgalmas írásban mutatja be a halhatatlan szubsztanciaként felfogott lélek beilleszthetlenségét a kritikai Kant gondolatvilágába. S noha a halhatatlanság teológiai vonatkozása el is halványul Varga Péter András írásában (*A husserli tudatfelfogás önértelmezési vitájának előképe a 19. században*), azért a pszichológizáló, valamely szubsztantív lélekfogalomnak legalább a maradványát Husserl íásaiban fellelni próbáló olvasatok felidézése, Husserl ütköztetése a korabeli pszichológizáló gondolkodásmóddal ehhez a gondolatkörhöz kapcsolódik. A politikafilozófia aspektusában a disztributív hatalomfelfogás oldalán áll Codruța Cuceu írásának nyilvánosságlelmzése (*A politikafilozófia modernizálása – A nyilvánosság pragmatista megközelítése*), míg Neumer Katalin annak jár utána, hogy a pszichológizáló transzcendentálfilozófia tárgyat konstituáló lélekfogalma helyébe lépő wittgensteini nyelvjáték milyen értelemben s milyen mértékben konstituálja a tényeket, illetve van-e „harmadik Witt-

genstein” (*A nyelvjátékok konstituálják-e a tényeket? Wittgenstein 1946–1951*). Vajda Mihálnál (*A lét hangoltsága*) az érzelmeket, affektusokat filozófiai-ontológiai szinten újra tematizáló heideggeri gondolkodás a szubsztantív lélek vagy elme (*res cogitans*) és a tőle élesen elválasztott, szubsztantív módon felfogott világ (*res extensa*) közötti túl élesen megvont határ feloldásának szükségességét nyomatékosítja. Elemzésében az unalom heideggeri hármassága áll a középpontban, amely végül összekapcsolódik a „szórakozás” pascali fogalmával.

*A Történelemre vonatkoztatott és politikumba átültetett áldozathozatali etika* című nagyobb blokk csaknem teljesen Lukács György gondolkodását vizsgálja – Hévi-zi Ottó a korai (*Személyességek, valóságok, etikák – A szókratikus Lukácsról*), Gángó Gábor a kései Lukácsal (*Sigfried Kracauer- és Walter Benjamin-hivatkozások Lukács György Az esztétikum sajátosságában*) foglalkozik –, illetve a Lukács-iskola és Lukács-óvoda tagjainak a marxizmustól egyre távolabb kerülő gondolkodási horizontját vázolja fel az *Überhaupt*-könyvtől kezdve (Kovács Gábor: *Az újbaloldal és a magyar filozófia a hervenes évek elején: néhány adalék az Überhaupt-könyv esztétörténeti háttéréhez*) a mára igen aktuálissá váló ökológiai gondolkodás magyarországi kezdeményéig Bence György és Kis János tanulmányában (Békes Vera: *Egy kézirat az ezerkilencszázhetvenes évek elején, a „márius és mégüres pillanat senkiföldjén”*). Fejtő Ferenc megidézésével ismét kitágul a magyar filozófia kifejezés jelentéstartományára (Agárdi Péter: *Fejtő Ferenc és a magyar nemzeti kultúra*), míg Ferencz Sándor és Surányi László írásával két olyan 20. századi gondolkodói kör bukkan fel, amelyek joggal tekinthetők a lukácsi, Marxból kiinduló filozofálás alternatíváinak. Ferencz Vidrány Katalin személyének és az általa művelt, a patrisztikára épülő teológiai-filozófiai gondolkodásnak állít



emléket („*Én nem az ő eretnekük vagyok*”), Surányi pedig Szabó Lajos munkásságához kapcsolódva adja a zene egyfajta filozófiai megközelítését (*A zene helye a dialogikus gondolkodásban*).

A kötet utolsó blokkja – *Miért bízunk inkább a tudósokban, mint a mágusokban?* – ismét a hatalomdisztribúció szálát bogozza tovább egy sajátos területen, a tudomány-filozófia és a kommunikációfilozófia területén. Laki János az egy meghatározott és mindenható tudományos módszerrel kutató és feltétlen bizonyosságra törekvő tudós helyébe az ilyen nagy igényeket nem támasztó tudományos közösségeket állítja, párhuzamot vonva a tudományos kiválóság és az etikai erényesség között (*Demarkáció: a megbízhatóság episztemológiája*). Kondor Zsuzsanna (*Túl a virtualitáson: mozgás, kép, szöveg*), Golden Dániel (*Az elektronikus olvasás mintázatai*) és Szécsi Gábor (*A közösség mint bizalmi folyamat. Közösségfogalom az elektronikus kommunikáció korában*) az egymással hálózatok révén összekapcsolódó, tehát individuumkénti elszigetelődésük-

ből kilépő s így *communitast* alkotó – kommunikáló – egyének világáról, virtualitásról és realitásról, az olvasás szétszorttá válásáról adnak filozófiai elemzéseket.

A kötet egésze az én perspektívámból – függetlenül az egyes írások eltéréseitől jellegükben, mélységükben stb. – egyfajta választ kínál arra a kérdésre, hogyan képzelhető el egy filozófiai intézet mint akadémiai kutatóhely működése ma, amikor az Archívum s az Intézet létrehozásának eredeti céljai – a „reformmarxizmus” művelése Lukács nyomán, Lukácsot kiigazítva – megszűntek vagy elhalványultak. A kötet egy olyan kutatói közösséget mutat be, amely képes elsajátítani és alkotóan továbbfejleszteni a nemzetközi filozófiai közösség kortárs gondolkodási irányzatait, képes ezeket a mi saját, magyar-kelet-közép-európai perspektívánkba ültetni, és így adni sajátos hozzájárulást a kérdések továbbgondolásához. Feltétlenül megőrzendő, a maga sokszínűségében ápolandó érték ez az egyetemek filozófiai műhelyei mellett, velük szoros együttműködésben.

## E számunk szerzői

Bene László (1968) az Eötvös Loránd Tudományegyetem BTK Filozófia Intézetének oktatója.

Kutatási területe az antik filozófia és recepciótörténete. E-mail: benel@elte.hu

Betegh Gábor (1968) a Közép-európai Egyetem (CEU) Filozófia Tanszékének oktatója. Kutatási területe az antik filozófia. E-mail: beteghg@ceu.hu

Boros Gábor tanszékvezető egyetemi tanár az ELTE BTK Újkori és Jelenkori Filozófia Tanszékén. Fő kutatási területe a 17–18. századi filozófiatörténet és a kora újkori szennvedély-elméletek. E-mail: gboros@ludens.elte.hu

E. Szabó László az Eötvös Loránd Tudományegyetem BTK Logika Tanszékének egyetemi tanára, a Logika és Tudományfilozófia doktori program vezetője. Fő kutatási területe: analitikus tudományfilozófia, ezen belül a fizika és a matematika filozófiai kérdései. E-mail: leszabo@phil.elte.hu

Farkas Katalin (DSc) a Közép-európai Egyetem (CEU) Filozófia Tanszékén tanít, a Közép-európai Egyetem tudományos rektorhelyettese. Főbb kutatási területei a kortárs elemfilozófia, a metafizika és Descartes. E-mail: farkask@ceu.hu

Forrai Gábor egyetemi tanár a Miskolci Egyetem Filozófia Intézetében. Főleg az ismeretelmélet, a metafizika és a nyelvfilozófia határterületére eső kérdésekkel foglalkozik. E-mail: forrai.gabor@upcmail.hu

- Gyenis Balázs PhD hallgató és részidős oktató a *University of Pittsburgh* Tudománytörténet és Tudományfilozófia Tanszékén. Fő kutatási területe: a formalizált tudományok, elsősorban a fizika filozófiai kérdései. E-mail: gyepi@hps.elte.hu
- Gyenis Zsolt matematikus PhD hallgató a Közép-európai Egyetemen (CEU). Fő kutatási területei a logika és a modellelmélet. E-mail: gyz@renyi.hu
- Huoranszki Ferenc (DSc) egyetemi tanár a Közép-európai Egyetem (CEU) Filozófia Tanszékén és az Eötvös Loránd Tudományegyetem BTK Filozófiai Intézetében. Főbb kutatási területei a kortárs metafizika és cselekvésemélet, valamint ezek kapcsolata a korai modern filozófiákkal. E-mail: huoransz@ceu.hu
- Márton Miklós (PhD) egyetemi tanársegéd az Eötvös Loránd Tudományegyetem Jogi Karának Filozófia Tanszékén. Főbb kutatási területei a kortárs nyelvfilozófia és észlelésemélet, illetve Hegel analitikus értelmezései. E-mail: marton@ajk.elte.hu
- Rédei Miklós egyetemi docens a *London School of Economics and Political Science, Philosophy, Logic and Scientific Method* Tanszékén, a *Philosophy of Science* MA Program vezetője. Fő kutatási területe az algebrai kvantumelmélet és a relativisztikus kvantumtérelmélet értelmezési és filozófiai kérdései. E-mail: m.redei@lse.ac.uk
- Simon József a Szegedi Tudományegyetem Filozófia Tanszékének adjunktusa. Fő kutatási területe a skolasztika és a korai modernitás filozófiai, ezek kapcsolódási pontjai, valamint a magyar filozófia története 1800-ig. E-mail: jsimon101@hotmail.com
- Szabó Gábor docens a Zsigmond Király Főiskola Filozófia és Vallástudományi Tanszékén, valamint Bolyai-ösztöndíjas az Eötvös Loránd Tudományegyetem BTK Logika Tanszékén. Fő kutatási területe: a fizika, elsősorban a kvantumelmélet, a valószínűség és a kauzalitás filozófiai problémái. E-mail: gsz@szig.hu
- Szalai Judit 2006-ban szerzett doktori fokozatot a Közép-európai Egyetem (CEU) Filozófia Tanszékén. Jelenleg az Eötvös Loránd Tudományegyetem tudományos segédmunkatársa. E-mail: jszalai@webmail.phil-inst.hu

## Summaries

GÁBOR BETEGH

'The Next Cause' – *Metaphysics* A 3–4

The paper is an abridged Hungarian version of the paper presented at the 18th Symposium Aristotelicum and which is forthcoming in the volume edited by Carlos Steel. On the basis of a close reading of the relevant sections of *Metaphysics* A 3–4 (984<sup>b</sup>8–985<sup>a</sup>29), it examines the way Aristotle describes the reasons and motivations, which, on his interpretation, lead his predecessors to introduce a second type of principle that could function as 'the source of change', i.e. the efficient cause. By bringing in parallel texts from the first book of the *Physics* and the first book of the *On the Parts of Animals*, the paper argues that, for Aristotle, the trajectory of the discovery of the truth is after all far less deterministic than what the language of *Metaphysics* A 3 might suggest. The paper aims to show, moreover, that what is discovered is not so much new types of Aristotelian causes but rather distinctions among types of principle. For, as Aristotle makes clear, already the Ionian monists used quasi-efficient causes by introducing opposites in the material substrate (hot–cold, fire–other elements). What is new in the thought of Anaxagoras and Empedocles is a new kind of principle, distinct from the material substrate that can assume the role of the efficient cause; but in part also that of the final cause. Finally, the paper tries to show why Aristotle's method of interpreting his predecessors within the framework of his own theory of four causes is not illegitimate.

## LÁSZLÓ BENE

## Causality and Moral Responsibility in Plotinus

In this paper, I examine Plotinus' view of the causal order of the physical world and his theory of moral responsibility.

(1) I highlight three characteristic features of Plotinus' causal theory. First, he accepts the general causal principle in relation to the cosmos („everything that comes to be has a cause”). Secondly, he introduces incorporeal, psychic causes in order to explain the generation and motion of perceptible things. Thirdly, he insists on the plurality of such independent causal principles: individual souls are not causally dependent on the world-soul, which is responsible for the structure and motion of the universe. I argue that Plotinus' theory of the causal order of the sensible world is, with some important qualifications, rather like Stoic determinism. His treatment of divination reveals that he rejects the idea of a genuinely open future. Further, he argues that providence takes into account individual human choices in advance in such a way that the latter cannot circumvent the providential plan. Finally, it is not the case that Plotinus introduces an indeterministic element into his causal theory by granting the individual soul the status of a principle. Unlike Middle Platonists, he does not refer to the possible (*dunaton*) or the contingent (*endechomenon*) in connection with moral responsibility. He does not separate character and choice – we are responsible for our choices precisely because they express the individual profile of our mind.

(2) For Plotinus, moral responsibility is grounded in the autonomy of the agent rather than in freedom to do otherwise. This kind of theory works only if it can be shown that we are responsible for our character. The Plotinian soul is not merely a product of its principle or of external causal influences. Metaphysically speaking, the soul is a self-constituting entity; furthermore, the actual state of discursive reason is a result of its acts in its previous life (Plotinus' account combines Platonic myths of transmigration with Aristotle's theory of moral development).

(3) Plotinus systematically connects moral responsibility with normative autonomy. Moral responsibility is grounded in the fact that the soul is a primary source of causation. In turn, the rational soul enjoys this causal status because it qualifies, at least in its original intellectual nature, as a “free principle”. In this way, freedom, taken in the sense of complete independence from external factors, is not only a normative ideal to be achieved, but it also has a major role to play in the explanation of responsible action. It has been claimed that normative freedom (*eleutheria*) and the compatibility of causal determinism with moral responsibility remain separate issues in Hellenistic and Imperial times. Plotinus' conception shows that this thesis needs to be qualified.

## JÓZSEF SIMON

## From the Causality of Intelligibility to the Intelligibility of Causality – Causality in Duns Scotus' Philosophy

Through a close reading of extracts taken from the *Liber de causis*, from Thomas Aquinas' Commentary on Boëthius' *De Trinitate* and from several works written by Duns Scotus, the paper examines how the role of causality changed from an explanatory principle of the intelligible properties of things into a mere characteristic of natural phenomena. It claims that Duns Scotus denied the causal efficacy of intelligible entities – even that of God – which served as an explanation for the ontological structure of things. For Scotus, this denial was possible through his innovative concept of possible intelligibility, which provided him with an ontology that required no causal explanation whatsoever. The paper

examines the characteristics of the so-called 'essential causal processes', which Scotus discusses in his reductive concept of causality when providing proof for God's existence as the first efficient cause.

JUDIT SZALAI

### Aristotelian-Scholastic Elements in Descartes' Theory of the Passions

This paper analyzes Descartes' account(s) of psychophysical interaction involved in the passions. According to its main argument, in certain respects, Descartes failed to abandon the Aristotelian-scholastic paradigm of sensation he criticized in a tendentious and unjustified way. This recognition does not render the Cartesian account of sensation sufficiently comprehensible, let alone plausible. However, it does highlight a factor that feeds into the problematic character of the Cartesian conception: Descartes' borrowings from the Aristotelian-scholastic tradition sit ill with some elements of his own philosophy of mind.

LÁSZLÓ E. SZABÓ – BALÁZS GYENIS –  
ZALÁN GYENIS – MIKLÓS RÉDEI – GÁBOR HOFER-SZABÓ  
Causal Explanation of Correlations

*No correlation without causation.* This is, in its most compact and general formulation, the essence of what has become called Reichenbach's Common Cause Principle. More explicitly the Common Cause Principle says that every correlation is either due to a direct causal effect linking the correlated entities, or is brought about by a third factor, the so-called common cause. Thus the Common Cause Principle is a non-trivial metaphysical claim about the causal structure of the World. In the essay the authors intend to give a brief overview of their fifteen year long research project devoted to the investigation of the status of the Common Cause Principle, or in other words, to the question of common causal explanation of correlations.

FERENC HUORANSZKI  
Causation at a Distance

The paper investigates the issue concerning the transitivity of causal relations. First, it claims that the assumption of transitivity is basic in the sense that it plays an essential role both in the explanation of causal asymmetry and in the debates about the possibility of reflective causation. Second, it shows why Humean, and more specifically, counterfactual accounts of causation are unable properly to explain when and why causation is transitive. Finally, it argues that transitivity is best explained by the agency-theory of causation. According to the agency-theory, an event cannot be a cause of another unless it is, or would be, rational for a free agent to choose the former as a means to produce the latter. Thus we might be able to understand transitivity from the perspective of the effects. Given means-ends relation, the events in the chain can be viewed as connected by 'by-relation', so that they constitute a chain of effective means to produce the final outcome. It is then the transitivity of 'by-relations' that explains why certain causal chains are also transitive processes.

A MAGYAR FILOZÓFIAI SZEMLE KÖVETKEZŐ SZÁMA:

*Imre Ruzsa: A Man of Consequence* (2010/4)

JÖVŐ ÉVI TEMATIKUS SZÁMAINK ÉS A KÉZIRATOK LEADÁSÁNAK HATÁRIDEJE:

*Irracionalitás* (2011/1) január 31.

*Hatalom* (2011/2) május 15.

*Tapasztalat* (2011/3) szeptember 15.

## SZERZŐINK FIGYELMÉBE

A *Magyar Filozófiai Szemle* referált folyóirat, amely csak eredeti kéziratokat közöl, és minden beérkezett tanulmányt két független, anonim bírálónak ad ki. A kéziratokat elektronikus formában várjuk a [filoszemle@webmail.phil-inst.hu](mailto:filoszemle@webmail.phil-inst.hu) címre. A teljes példány mellé egy másik, név és önhivatkozások nélküli változatot is kérünk a referensek részére.

A kéziratokat Times New Roman betűtípusban, „rtf” formátumban, körülbelül egyívnyi (40 000 leütés) terjedelemben, csupán a legszükségesebb formázásokkal (például kurzíválás) kérjük. A főszövegben a 4-5 sornál terjedelmesebb idézetek a törzsszövegtől egy-egy sorral elválasztva, behúzva szerepeljenek.

Tanulmányok esetében a hivatkozott szakirodalmat a szövegbe illesztve kérjük feltüntetni: (Kuhn 1999a. 99), vagy lábjegyzetben: Putnam 1975b. 237. Szükség esetén megadható az eredeti megjelenés dátuma és a kötetszám is (Descartes 1644/1996. 8/2:72). Használhatók a széles körben elterjedt rövidítések (AT VIII-2. 72), és klasszikus szövegek esetében a bevett utalási eljárások (Platón, *Phaidrosz* XXIV 245c–d; Mt 9,13). A rövidítések feloldását kérjük az irások végén, az irodalomjegyzekben megadni.

Az irodalomjegyzéket a következő példákhoz megfelelően kérjük összeállítani:

Önálló kötet:

Locke, John 1690/2003. *Értekezés az emberi értelemről*. Ford. Vassányi Miklós – Csordás David. Budapest, Osiris.

Tanulmány gyűjteményes kötetben:

Henke, Christopher R. – Thomas F. Gieryn 2008. Sites of Scientific Practice: The Enduring Importance of Place. In Edward J. Hackett – Olga Amsterdamska – Michael Lynch – Judy Wajcman (szerk.) *The Handbook of Science and Technology Studies*. Cambridge/MA, The MIT Press. 354–376.

Periodikumban megjelent cikk, tanulmány:

Schnadelbach, Herbert 1993. A filozófia és az embertudományok. Ford. Mezei György Iván. *Magyar Filozófiai Szemle*, 37/2. 1015–1031.

A kéziratokhoz rövid bemutatkozást és angol nyelvű összefoglalót kérünk mellékelni

# MAGYAR FILOZÓFIAI SZEMLE

2010/3 (54. évfolyam)

„Íme egy asztal, s rajta két biliárdgolyó: az egyik áll, a másik gyorsan közeledik feléje.” Talán nincs egyszerűbb és ismerősebb tárgya a filozófiai vizsgálódásoknak, mint az ok–okozati viszony, amely egyszerre alapja a mindennapi életben való tájékozódásnak (miként azt Hume hétköznapi példája sugallja), s a különböző fizikai vagy társadalmi jelenségekkel kapcsolatos tudományos elméletalkotásnak, hiszen – amint ugyancsak Hume állítja – „a tényeket illető következtetéseink mindig az ok és okozat relációján alapulnak, és [...] soha nem tudunk az egyik tárgynak a létezéséből a másikéra következtetni, ha a két tárgy nem áll közvetett, vagy közvetlen kapcsolatban egymással”. És mégis, az okság ismerős jellege dacára máig nyitott kérdés, hogy miben is áll két esemény kauzális viszonya. Elemezhető-e más metafizikai fogalmak segítségével, vagy redukálhatatlan reláció? Melyek e reláció terminusai? Tárgyak, események, vagy tények közötti viszony-e, esetleg egyéb relátumokkal rendelkezik? Helyes-e Hume intuíciója, amely szerint az okság csupán regularitás; vagy létezik szinguláris kauzalitás is, netán a kauzalitás az emberi cselekvésre épül?

ISSN 0025-0090



1200 Ft

GONDOLAT

**nka**  
Nemzeti Kulturális Alap

P 25591

Ps2



2011 FFRR U 8

A Magyar Tudományos Akadémia  
Filozófiai Bizottságának folyóirata

# HUNGARIAN PHILOSOPHICAL REVIEW

2010/4 (54. évfolyam)

Imre Ruzsa—A Man of Consequence

## MAGYAR FILOZÓFIAI SZEMLE



## SZERKESZTŐSÉG

Schmal Dániel felelős szerkesztő, Pázmány Péter Katolikus Egyetem  
Ambrus Gergely, Miskolci Egyetem  
Mekis Péter, Eötvös Loránd Tudományegyetem  
Mester Béla, MTA Filozófiai Kutatóintézet  
Ullmann Tamás, Eötvös Loránd Tudományegyetem  
Valastyán Tamás, Debreceni Egyetem

## SZERKESZTŐBIZOTTSÁG

Forrai Gábor, a szerkesztőbizottság elnöke, Miskolci Egyetem  
Almási Miklós, az MTA rendes tagja, Eötvös Loránd Tudományegyetem  
Altrichter Ferenc, University of North Caroline at Wilmington  
Csejtei Dezső, Szegedi Tudományegyetem  
Fehér Márta, Budapesti Műszaki és Gazdaságtudományi Egyetem  
Gendler Szabó Zoltán, Yale University  
Heller Ágnes, az MTA rendes tagja, Eötvös Loránd Tudományegyetem, The New School for Social Research (New York)  
Hévízi Ottó, MTA Filozófiai Kutatóintézet Lukács Archívum  
Hont István, King's College (Cambridge)  
Kelemen János, az MTA levelező tagja, Eötvös Loránd Tudományegyetem  
Klima Gyula, Fordham University (New York)  
Máté András, Eötvös Loránd Tudományegyetem  
Rózsa Erzsébet, Debreceni Egyetem  
Tengelyi László, Bergische Universität (Wuppertal)  
Weiss János, Pécsi Tudományegyetem

A szerkesztőség elektronikus postacíme:  
filszemle@webmail.phil-inst.hu  
schmal.daniel@chello.hu

ISSN 0025-0090

Kiadja a Gondolat Kiadó  
Felelős kiadó Bácskai István  
1088 Budapest, Szentkirályi u. 16.  
Telefon: + 361-486-1527  
www.gondolatkiado.hu  
http://gondolatkiado.blog.hu  
A lap kiadásában közreműködött a BÉGé Bt.

Tördelő: Lipót Éva  
Lapterv: Andorka Tímea és Pintér László  
Nyomás és kötés: Nyomdai munkák  
*mondAt Kft.* • Budapest  
www.mondat.hu

A lap megjelenését a Nemzeti Kulturális Alap  
és a Magyar Tudományos Akadémia Filozófiai Bizottsága támogatta



# HUNGARIAN PHILOSOPHICAL REVIEW



# MAGYAR FILOZÓFIAI SZEMLE

2010/4 (54. évfolyam)

---

A Magyar Tudományos Akadémia  
Filozófiai Bizottságának folyóirata

Imre Ruzsa—  
A Man of Consequence



# Contents

Foreword	5
----------	---

## **RUZSA'S WORK**

ANDRÁS MÁTÉ, Imre Ruzsa—A Man of Consequence	7
FERENC CSABA, Whose Logic Is Three-Valued Logic?	19
TAMÁS MIHÁLYDEÁK, On Models of General Type-Theoretical Languages	27
ZSÓFIA ZVOLENSZKY, Ruzsa on Quine's Argument against Modal Logic	40

## **PHILOSOPHICAL LOGIC AND ITS HISTORY**

ANNA BROŽEK, On the So-Called Embedded Questions	49
GYULA KLIMA, Natural Logic, Medieval Logic and Formal Semantics	58
EDWARD KANTERIAN, Frege's Definition of Number: No Ontological Agenda?	76
NENAD MISCEVIĆ, The Indispensability of Logic	93
EDI PAVLOVIĆ, Fitch's Paradox and Labeled Natural Deduction System	104
JÍŘI RAČLAVSKÝ, On Partiality and Tichý's Transparent Intensional Logic	120
MÁRTA UJVÁRI, Prior on Radical Coming-into-being	129

## **FORMAL SEMANTICS**

LÁSZLÓ KÁLMÁN, Analogy in Semantics	134
ANDRÁS KORNAI, The Treatment of Ordinary Quantification in English Proper	150
PÉTER MEKIS, Atomic Descriptions in Dynamic Predicate Logic	163

## **PHILOSOPHY OF MATHEMATICS**

ZOLTÁN GENDLER SZABÓ, Tasks and Ultra-tasks	177
GÁBOR FORRAI, What Mathematicians Say Means: In Defense of Hermeneutic Fictionalism	191

## **FOUNDATIONS OF SCIENCE**

HAJNAL ANDRÉKA—JUDIT MADARÁSZ-- ISTVÁN NÉMETI—GERGELY SZÉKELY, On Logical Analysis of Relativity Theories	204
ROBIN HIRSCH, Modal Logic and Relativity	223
MÁTÉ SZABÓ, On Field's Nominalization of Physical Theories	231
Contributors	240

## Foreword

This special issue of the Hungarian Philosophical Review is dedicated to the memory of our friend, colleague and teacher Imre Ruzsa (1921–2008). Ruzsa was the father of modern philosophical logic in Hungary, the founder and the first chair of the Department of Logic at Eötvös University. His professional interests centered around modal logic, intensional logic, modeling natural language in systems of intensional logic, and the foundations of logic and mathematics. He always thought of his generalization of A. N. Prior's concept of semantic value gaps to quantified, intensional and type-theoretic systems as his most important contribution to logic. He was the author of three books in English (*Modal Logic with Descriptions*, The Hague, 1982, *Intensional Logic Revisited*, Budapest, 1991, *Introduction to Metalogic*, Budapest, 1993), several monographs and textbooks in Hungarian, and many articles in leading logic journals.

The variety of the papers included in this volume represents the range of topics that Ruzsa's research covered: philosophical logic, formal semantics, the philosophy of mathematics, and foundational studies. The papers were all given at the Imre Ruzsa Memorial Conference "Logic, Language, Mathematics", held at the Philosophy Institute of Eötvös University in Budapest on September 17–19, 2009, as part of the annual conference series "Language, Understanding, Interpretation."<sup>1</sup> It is worth mentioning the invited speakers who gave memorable presentations that are published elsewhere: Rob Goldblatt (Victoria University, Wellington), Ági Kurucz (King's College London), Mihály Makkai (McGill University, Toronto), László Pólos (University of Durham, UK), and Anna Szabolcsi (New York University).

Among the authors the reader will find three generations of logicians: fellow researchers whose work Ruzsa's ideas influenced strongly; former students of Ruzsa, who had once been introduced into the mysteries of the logical connectives in seminar rooms of Eötvös University, and many of whom are today lead-

<sup>1</sup> For more information about the conference, see:  
<http://phil.elte.hu/ruzsaconf>.

ing researchers in their field, teaching across the globe; and a younger generation of logicians, Ruzsa's "grandchildren", whose work testifies to the enduring influence of Ruzsa's legacy.

The editors would like to thank all the authors for their contributions, among them Zsófia Zvolenszky, whose language editing work has also been invaluable.

*András Máté and Péter Mekis*

ANDRÁS MÁTÉ

## Imre Ruzsa—A Man of Consequence

**Abstract.** The singular aim and task of this paper is to present an overview of the life and work of Imre Ruzsa.

### 1 AN UNUSUAL ROUTE TO PHILOSOPHY

Ruzsa's life was rather different from a typical academic career. He was born on the 12<sup>th</sup> of May in 1921 in Budapest but grew up in a little town in the south-eastern part of Hungary as the son of a tailor. His family couldn't send him to high school, so after finishing elementary school, he worked as an assistant to his father. At the age of seventeen, he left his father's house and worked as a tailor's assistant first in Debrecen, a city in Eastern Hungary, then in Budapest. He joined to the Social Democratic Party once in Budapest, and was admitted to the illegal Communist Party. He worked as the printer for the party newspaper and for this activity, he was convicted to eleven years in prison in 1942. The first time he heard about mathematics beyond the common arithmetical operations was from a Communist economist in the prison courtyard. Like other political prisoners, he was sent to the front lines in a forced labour company in 1944. He escaped and survived the rest of the war with false documents in Budapest.

After the war, he finished high school on an accelerated track, then in 1947, he began his university studies at the Faculty of the Humanities of Budapest University. He attended diverse lectures on philosophy and Hungarian linguistics. After the reorganization of university programs in 1950, he became a student of mathematics, physics and descriptive geometry at the new Faculty of Science. By taking a look at his university records, we can note the intellectual level of the Mathematical Institute of Budapest University in those years: in fact, to this day, all professors listed there are regarded as prominent figures of the history of modern mathematics. At the very least, let me mention the name of Lipót Fejér, Rózsa Péter and Alfréd Rényi. Ruzsa's academic records indicate that he fulfilled the requirements of this institute with flying colors. In personal

conversations, he himself remarked that his mathematical studies allowed him to keep a bit more of a distance from politics at that point. But politics and history didn't release him: in the year 1953, a few weeks before Stalin's death, he was arrested again and sentenced to five years for "war crimes and crimes against the people". He was set free in the next year and rehabilitated in 1957.

In 1956, he finished his university studies and began teaching mathematics in a geological polytechnic. In 1960, he was invited to the University as a lecturer of mathematical calculus. From 1962 on, he taught mathematics for philosophy students. In these years, he began his research into modal and deontic logic. Earlier, while still a student, he had studied mathematical logic with Rózsa Péter, by this time, they were at the same department—Rózsa Péter read Ruzsa's writings in the sixties and gave him extensive advice and comments, especially on the philosophy of mathematics and on the mathematical aspects of logic. Ruzsa was in contact with the other great master of mathematical logic in Hungary, László Kalmár, too, but in philosophical logic, Ruzsa had no ancestors and mentors in Hungary at all, nor did he have any opportunities to study abroad either. He simply used the library and began corresponding with Arthur N. Prior, whose ideas had the greatest influence on him.

## 2 EARLY WORK IN PHILOSOPHY

In the year 1965, the Philosophy Institute of the Hungarian Academy of Sciences offered Ruzsa a job as a research fellow. He accepted it but kept his position at the university as a part-time assistant professor. During the sixties, he wrote several pieces about the philosophy of mathematics: a book for teachers (1967), an article in the *Hungarian Philosophical Review*, a series of articles for scientists and science teachers, remarkable lecture notes on mathematics for students of philosophy (1964), as well as a book for the broader public (1968). Ruzsa had many popular writings on mathematics and often connected the popularization of mathematics with philosophy. He was awarded the Manó Beke prize for popularizing mathematics in 1971.

A second group of his early papers consists of eight survey papers about contemporary research in philosophical logic for the *Hungarian Philosophical Review*. It was not merely academic reasons that led him to write the last three of these about research on symbolic logic in the Soviet Union. Ruzsa needed to prove that modern logic in philosophy didn't threaten the ideological foundations of the Communist regime. He presented what was effectively an argument from authority, showing that symbolic logic was an accepted research area in the Soviet Union. Investigations by Soviet logicians at that time, like Vladimir Smirnov or Aleksandr Zinoviev, who was later to become a political dissident, and some others, were carried out in accordance with logical research at leading



Western universities. Within Hungarian philosophy, during the sixties, Ruzsa stood almost alone with his research program. It would be unjust to deny that there were some philosophers who tried to integrate some tools and ideas of modern logic into university education and research; but mostly these efforts led to no more than a confused mixture of modern and obsolete ideas. In the realm of education, it was Sándor Szalai who did the best work in terms of integrating some modern logic into the logic curriculum at Budapest University during the forties and fifties; but the effects of his work were limited because Szalai was not a logician, not even a philosopher, but a sociologist who knew a fair amount about logic. The situation was paradoxical because in mathematical logic, Hungary had an abundance of great scholars, such as László Kalmár, Rózsa Péter and their numerous students. It was Ruzsa's mission to convey their knowledge to philosophy.

The third group of papers includes the first results of his own research in logic. He focused on two topics: deontic logic and the connection between logic and probability theory. His doctoral thesis at the Hungarian Academy of Sciences (the degree was called "candidate of mathematical science") was based on the latter topic. The title was "Random models of logical systems"; it was prepared without an official supervisor and Ruzsa mentions in the documents no mentor or advisor in Hungary. Ruzsa didn't subsequently return to this topic. Another branch of his early work was, however, the very beginning of a continuous line of research for the decades to come. Deontic systems are in fact special cases of modal logic; Ruzsa's survey papers from the same time display his interest into general modal logic and Kripke semantics. The idea of semantic value gaps which turned out to be the central thought of his logical work emerged during this time from his study of Arthur Prior's work and from correspondence with him.

### 3 AT THE DEPARTMENT OF LOGIC

In 1970, both the structure of the departments and the curriculum for philosophy students was reorganised at Eötvös University in Budapest. Mathematics was banned from the curriculum, but Ruzsa received a new task: he joined in the teaching of logic. The newly founded Department of Logic was in charge of the course of study in logic, which consisted of two main components up until the transition period in 1989-90: two or three semesters of formal logic and two semesters of dialectical logic. The basic principle was that the true logic of Marxist-Leninist philosophy was dialectical logic and formal logic was just a subordinated preliminary study to it. Dialectical logic meant, according to the head of department, a sort of materialistically transformed Hegelian logic; it in fact required no more formal logic than a minimal knowledge of Aristotelian syllogistic.

However, this situation made it possible to teach some real logic to young philosophers as long as one resigned oneself to steering clear of questioning the superiority of dialectical logic. Ruzsa published the first version of his lecture notes in logic in 1969, even though he taught the lectures on logic only the next year, and accepted the invitation to the Department of Logic as an associated professor in 1971. This was a great turn both in Ruzsa's life and in logic education.

There was a threefold difference between earlier "formal" logicians at Budapest university and Ruzsa. Firstly, he had the requisite mathematical background to follow contemporary research and contribute to it. Secondly, he didn't bother with improving and modernising old teaching materials and curricula but wrote a completely new one built on modern logic (and improved it over the next thirty years).<sup>1</sup> Thirdly, he didn't go into discussions about what real dialectical logic was supposed to be. Other people in Hungary, as well as in other Eastern-block countries, tried to sell under the name "dialectical logic" some more or less modern methodology of science and were therefore drawn into conflicts with Hegelian dialectical logicians. Ruzsa didn't interfere with the affairs of dialectical logicians. Instead, he responded in sarcastical short articles when mathematical logic was attacked for sneaking antidialectic, metaphysical, neopositivistic etc. ways of thinking into Marxist philosophy, charges brought on by people who had no real knowledge about the subject. As he was a fellow at a department led by the most militant dialectical logician, he didn't expect anything more in those years than that they leave him to work and teach.

In spite of the often astonishing circumstances, the seventies were fruitful years for Ruzsa both in terms of research and teaching. In modal logic, he generalized Prior's idea of truth value gaps to semantic value gaps and on this basis, he elaborated a Kripke-style semantics for various systems of first-order modal logic. His first paper about these systems was his (1973b). His dissertation based on this research, entitled *Individuals in modal logic*, earned him the degree "Doctor of Philosophical Science" at the Hungarian Academy of Sciences. The expanded English version of the dissertation was published by Martinus Nijhoff Publishers (1981). Ruzsa was appointed full professor in 1978.

In terms of teaching logic, beyond the lectures for philosophy students, Ruzsa was given the task of teaching logic and mathematics for students in theoretical linguistics. There was a lucky coincidence between this task and his new interest in the logical modeling of natural languages. Ruzsa became acquainted with Montague semantics in mid-seventies and immediately began to investigate how the idea of semantic value gaps might be implemented into Montague grammar. This idea led to more substantial changes in Montague semantics than

<sup>1</sup>Let us emphasize among the different versions the legendary "three-volume one" (1973).

in Kripke semantics but also proved to be even more fruitful.<sup>2</sup> For these investigations and for his teaching activity, he received recognition within a circle of younger linguists and some of them joined him as personal students, participants and guest speakers at his seminars in the seventies and eighties.

In 1977, an opportunity arose to expand the group of “formal logicians” within the Department of Logic with two new lecturer appointments. Ruzsa planned to orchestrate the celebration of the 100th birthday of symbolic logic (the centenary of Frege’s *Begriffsschrift*), and as a first task, he assigned to one of the new lecturers (namely, me) the translation of a selection from Frege’s writings. This volume was published with a slight delay, in 1980, accompanied by an issue of the Hungarian Philosophical Review which contained numerous papers on Frege and modern logic and translations of Frege’s articles “The Thought” and “Negation”. It was only Ruzsa’s own extensive programmatic article on Frege and the importance of modern logic for philosophy (1979) that was published exactly for the centenary in the Hungarian Philosophical Review. That is, the Review did not undertake to publish a special Frege-issue, as Ruzsa’s original intention had been. Nevertheless, the centenary of the *Begriffsschrift* was an important step towards the formation of Ruzsa’s school. The work of Frege offered a common starting point for the areas where Ruzsa’s writings and educational activity gained influence over the previous years: logic, philosophy of mathematics, linguistics, philosophy of language. The works published for this occasion therefore reached all the actual and potential students of Ruzsa and sympathizers of his work.

Moreover, this time, Ruzsa could appear in public together with some of his students and members of his circle. In Hungarian philosophy, symbolic logic had often been associated with logical positivism (not only in the era of Marxism-Leninism, but earlier, too). When Ruzsa, as a modern logician, was accused of smuggling neopositivistic influences into Marxist philosophy, he found this charge was awkward and at the same time, also amusing. For he had no special sympathy for logical positivism at all. Carnap belonged, of course, to the most widely cited authors in his monographs, but mostly it was not out of Ruzsa’s agreement with Carnap’s claims. Ruzsa was much more inclined towards realism; he was not a Frege-type Platonist, but his position was closer to Frege than to Carnap. This way, the centenary celebration also offered an opportunity to present as the founding father of symbolic logic a thinker as far from neo-positivism or any sort of positivism as Frege was.

<sup>2</sup> His first publication on this area was (1980).

## 4 THE NEW DEPARTMENT

Through the eighties, the Ruzsa school flourished. There was a favourable turn of circumstances: in 1982, the unwanted marriage with dialectical logic could be broken off and a new "Group for symbolic logic and the methodology of science" was founded, headed by Ruzsa. It was an odd, unconventional unit that was subordinated to no departments but to the institute of philosophy ("Marxism-Leninism") only; yet it didn't have the rank and rights of a department until 1984. The new department began to publish a yearbook called *Tertium non datur*. Besides the members of the department and Ruzsa's PhD students, several linguists, philosophers and other scholars wrote articles and reviews for the *Tertium*; its table of contents showed that Ruzsa and his circle were now gaining considerable influence in the humanities as well, among people interested in modern methodology.<sup>3</sup> In the yearbook we could now break with the earlier strategy of keeping distance from debates; by this time, it contained several sharply critical papers. In the opening volume, Ruzsa and five of his younger colleagues published a paper that dissected a logic textbook that was in use at teacher-training colleges and unified obsolete ideas of traditional logic with dialectic materialistic slogans. At times, the analysis would change into satire. We had considerable fun putting together this critique, but didn't foresee the consequences of it: during the next academic year, the textbook was withdrawn by the ministry of education. But to tell the truth, we didn't yet gather up the confidence to criticize dialectical logic.

Perhaps this is the time to say something about Imre himself as a man. It is not easy because his personality was rather hidden. Autonomy and steadfastness were among his major traits. He had chosen a path for himself and nobody could divert him from it. He wasn't interested in success, praise or money; he did what he thought was the right thing to do and that was all there was to it. He was very helpful. I think most of the colleagues who knew him are indebted to him, but only few of us can claim to have had a truly personal conversation with him. He endured the humiliating situations that occurred at the Department of Logic with calm irony and on rare occasions, with sarcastic remarks—it was only by the end of the nineties that I understood how deeply he was insulted by them, when we compiled a repertory volume from the volumes of *Tertium non datur* and he wanted to devote two pages of a four-page foreword to this topic. I needed hours to convince him that Comrade Erdei (the head of that department) didn't deserve so much attention any more. Well, his good sense of humour and irony

<sup>3</sup> Let me illustrate this influence by quoting the names of the Hungarian authors of *Tertium*: András Bárd, Katalin Bimbó, István Bodnár M., Balázs Dajka, Katalin É. Kiss, Özséb Horányi, Márta Fehér, László Kálmán, Ferenc Kiefer, Gyula Klima, Imre Komlósi L., András Kornai, Judit Maár, Anna Madarász Zsigmond, Márta Maleczki, András Máté, Tamás Mihálydeák, Sándor Ónódy, Kornél Solt, Anna Szabolesi, Zoltán Szabó [Gendler], Tibor Szécsényi.

often helped him in difficult situations. He was a quiet person, never a loud word even if he was angered. On the other side, during the occasional relaxed moment, he liked to make jokes. In the eighties, he wrote a “Dictionary for patho-logicians”. Its entries contain “explanations” of logical notions that often mix wordplay with pin-pricks at colleagues and profound remarks. Unfortunately, most of them are basically untranslatable wordplays in Hungarian, but let me quote one that (hopefully) works in English as well:

*Inconsistency*: a heavy and contagious disease. Especially widespread among philologists. The reduction of texts is the only cure.

In the eighties, Ruzsa published two large monographs. The first was *Classical, modal and intensional logic* (1984). It contained less technical details but a thorough analysis of the philosophical literature on logic, especially on the logic of modalities. In this book, Ruzsa explored the philosophical motivations behind his logic with semantic value gaps and gave in-depth arguments about its advantages. The second book is the two-volume *Logical Syntax and Semantics*—volume I: (1988), volume II: (1989), in which the author gives a self-contained, comprehensive introduction to logical theory, together with the foundations of logical syntax and completing his survey with a description of a formalized fragment of Hungarian. This is the main work of Ruzsa; I shall say more in a bit about its first, metalogical chapter.

## 5 THE LAST YEARS

Ruzsa retired from professorship and from chairing the department in 1990. This unavoidable step together with the fact that some members of the department along with other colleagues from the Ruzsa-circle went on to pursue their careers abroad, at outstanding universities—which was otherwise very much a welcome fact—made the activities of the department somewhat more difficult and less effective. On the other side, at the end of the eighties began our cooperation with the Algebraic Logic department of the Alfréd Rényi Mathematical Institute of the Academy, which made it possible to start the Logic Graduate School, one of the very first graduate programs in Hungary. Although formally Ruzsa was not the leader of this graduate school, he did play a definitive role in its first years, up until the end of the 1990s. He published improved English versions of the two most important chapters of *Logical syntax and semantics: Intensional logic revisited* (1991) and *Introduction to metalogic* (1997). In 1991 he was awarded the Széchenyi Prize, the highest state honour for achievements in science. In 1998 he was appointed professor emeritus. His last larger work was a new textbook of logic (1998), even more comprehensive than the three-volume one.

Ruzsa's advanced age and failing health gradually decreased his involvement in logic and the department. But his former students who visited him over the last years had the chance to witness his spirit remaining the same throughout.

## 6 PHILOSOPHY OF MATHEMATICS

After this biographical outline, let me speak in some detail about Ruzsa's work in two closely connected areas that received less attention in the conference program: philosophy of mathematics and metalogic. Through the sixties, his writings on the philosophy of mathematics emerged not so much from his research interests, but mostly as responses to the interests of his readers. As he writes in the foreword of *Between Mathematics and Philosophy*, he observed that many of his students couldn't buy the lecture notes (1964) to his mathematics lectures for philosophy students (in which he discussed foundations and philosophy of mathematics in detail) because the copies were bought off by interested outsiders. He had written a book presupposing some mathematical knowledge, mainly for teachers of mathematics, but it was, again, not enough. So he published *Between Mathematics and Philosophy* (1968) for the larger public, setting forth in a popularizing style the mathematical background needed. But he didn't regard this area his field of research; his goal was merely to summarize the basics of various trends in the philosophy of mathematics from his own perspective, and convey them to the Hungarian public. So we can't speak about his philosophy of mathematics in the proper sense, I will therefore content myself with characterizing Ruzsa's point of view.

Ruzsa focuses on introducing the three classical schools in the philosophy of mathematics: logicism, intuitionism and the Hilbert-school or formalism. His main stress is on the contents of and mathematical motivations behind these trends, their connection with research in the foundations of mathematics, but he also sets forth some critical remarks. His general opinion is that all the three schools capture something from the real nature of mathematics, but each of them is one-sided; that is, he argues for some sort of eclecticism. He has the most sympathy for Hilbert's program which he demarcates from the formalist philosophy of mathematics. He agrees with this program in that foundational problems should be solved by mathematical tools, while staying away from destroying what was constructed in mathematics. He argues that Gödel's second incompleteness theorem has serious consequences for Hilbert's program but he does not consider them fatal. But he does criticize formalists for rejecting the importance of content in mathematics. Theorems of mathematics have their content, they are true propositions and it happens only for the sake of metamathematical investigations that we abstract from their content and regard them just as syntactical strings. This moderate, realistic understanding of Hilbert's

program and the sympathy for it is characteristic of the philosophical writings of Péter and Kalmár, too. But logicism is evaluated by Ruzsa in a more favorable way than by his predecessors, as he lays more stress on the philosophical, realist side of Frege's and Russell's logicism. On the other side, he criticizes intuitionists rather sharply.

Why were these writings of Ruzsa that contain little by way of original insights so popular during the sixties? Today, readers may be astonished by the occasional Marxist detours and quotations from Engels or Lenin in these works. The official prescriptions of that era were such that it was allowed to expose non-Marxist philosophical views but only when the exposition was accompanied by a thorough Marxist criticism of them. Beyond the fact that Ruzsa surveyed an area that was virtually unknown in Hungary, the novelty of his writings was that he devoted far more space to the exposition and objective analysis of the various philosophies of mathematics than to their criticism. Actually, this was a similar approach as the one found in the works of the circle of George Lukács. For example, in the book *Trends in contemporary bourgeois philosophy* by György Márkus and Zádor Tordai from 1964, we find similar efforts: the authors present the different philosophical schools and thinkers from a Marxist perspective, but the primary stress is on the exposition and analysis of the views. There was a rather sizeable distance between this attitude and the practice of Soviet Marxism, whose main concern was to classify non-Marxist thinkers as mechanical materialists, objective and subjective idealists and to discover traces of Marxist truth in their writings. It must be remarked that the Marxist detours were sincere—both from the side of Ruzsa and from the Lukácsists. They all had some rather abstract commitment to Marxism and socialism—not the actual positions of party ideologists, of course. They tried to preserve as much from Marxism as they found acceptable—there was, of course, no place for criticizing Marxism where it was not acceptable.

## 7 METALOGIC AND THE PHILOSOPHY OF LOGIC

After 1970, Ruzsa stopped publishing on the philosophy of mathematics. However, in his main work *Logical Syntax and Semantics* (1988, 1989) he made an important contribution to the circularity problem in foundations, that is, to the problem that logic has its semantical foundations in set theory but on the other hand, set theory is a theory which can prove its theorems within a logical framework. Ruzsa had always taught that a logical theory is useless if it has no intuitively acceptable semantical foundations. He criticized relevant logics of Zinoviev and systems of entailment for lacking such foundations and regarded Kripke semantics not only as a technical tool but as a way to make explicit the real content of modal logic. In his textbooks and lecture notes, the language

of logic is introduced in a purely semantical way. Inference rules are just mentioned but hardly anything further than that; on the elementary level, there are no formal deductions at all. Students should learn how to check the validity of a given inference by the methods of truth-tables, Venn-diagrams or semantic tableaux; they are not expected to find out consequences of a given set of premises. The methods are semantical and their correctness is likewise confirmed by informal semantical considerations. (There is, of course, no formal semantics at this level.) This is in accord with Ruzsa's rather strong realist commitment that is present in his writings about the philosophy of mathematics and in the paragraphs and chapters concerning the philosophy of logic in his logical writings. The method to begin logic with semantically defined logical constants is present in *Logical Syntax and Semantics*, too; but quite surprising, that actually shows why Ruzsa was not a Platonist, in spite of all of his realist commitments.

The logical theory constructed there starts with introducing symbols of first-order logic—logical constants and variables—into the language of communication (metalanguage). The single difference between their introduction and the usual way is that every variable is declared, that is, it is specified what values they are allowed to take. In this way, the extended metalanguage preserves the property presupposed about the language of communication, namely, that every proposition has one and only one truth-value. Only this much is needed by way of informal semantical considerations behind the metalanguage logic. In order to prove that the axioms of this theory are true, the metalanguage is extended with class abstractions that are constructed from monadic open sentences and it is enough to introduce some minimal class theory which needs no axioms but just definitions of the empty set, the subset relation and the usual binary operations.

The concepts and assumptions needed for the theory of canonical calculi concern language as the class of expressions, that is, finite strings over a finite but nonempty alphabet (the class of letters). Using the operation of concatenation, the class of expressions can be described in an axiomatic way. Canonical calculi define inductive classes within the class of all expressions as strings deducible by a given finite set of rules. (Rules contain mostly a distinguished letter not contained in the alphabet: the arrow, and we are also allowed to use other auxiliary letters.) This very simple machinery suffices for the following:

- to represent calculi by strings of the original alphabet;
- to produce hypercalculi that define the class of all calculi;
- to introduce Gödel numbering, using as “numbers” the strings formed solely from an arbitrary element of the alphabet;
- to prove that there are certain subclasses of the language that can be defined in the metalanguage but are not inductive (although their complements with respect to the set of all expressions are).



This last claim is in fact a Gödel-type theorem now. The following step is the introduction of Markov-algorithms that is natural and easy in this language because the formalisms of canonical calculi and Markov-algorithms are very similar. Roughly speaking, the single difference between the two is that in executing an algorithm, the next step is always determined; in executing a deduction within a canonical calculus, we are free to choose the next step among the allowed ones. Enumerability and decidability by algorithms are defined as usual, and it is easy to show that enumerable sets of expressions are the same as the inductively definable ones. A set is decidable iff both the set itself and its complement is enumerable. With respect to these facts, the theorem mentioned above has as a simple corollary a Church-type theorem: there are enumerable but undecidable sets of expressions.

Real first-order logic follows only after this theory of canonical calculi and algorithms. We can inductively define the language of first-order logic and the set of provable formulas. Within this first-order logic, the theory of canonical calculi (CC) can be formalized and we can prove via metalanguage argumentation that all the theorems of CC are true. In fact, this is the only statement for which we need to use metalanguage logic and set theory. In other words, metalanguage logic has to be accepted on the basis of intuitive semantical background considerations only as far as it is applied to classes of expressions, that is, strictly finite objects. The only place for infinity is that we need a weak form of the induction principle in our metalanguage argumentation. Metalanguage set theory is basically no more than an inventory of abbreviations; its theorems are in fact truths of metalanguage logic.

Everything else turns out surprisingly simple: it follows from the previous theorems that CC is not decidable and every theory which is an inductive class of theorems containing CC is negation-incomplete. Real set theory is a first-order theory defined inductively, and we can use set theoretical propositions in constructing semantics for first-order logic only if we can prove them within this first-order set theory. This whole construction is the answer to the question of how the priority of semantics should be understood: we should accept some semantical considerations before we can construct the syntax of our logic, but these considerations are reduced to a minimum that fulfils the Hilbertian requirement of finiteness. In the formal construction, the priority belongs to syntax and deducibility; there is no Platonic heaven of mathematical objects that we know about without knowing an axiomatic theory of them. Most of the details of Ruzsa's construction of the foundations of logic are not his own inventions; but the construction as a whole is both well-considered and well-founded on the philosophical side and elegant on the mathematical side.<sup>4</sup>

<sup>4</sup> This paper was supported by the Hungarian National Scientific Research Foundation OTKA, project No. 68043.

## REFERENCES

All the items in Hungarian if not indicated otherwise.

- Ruzsa, Imre, 1964, *Mathematics for philosophy students*. Lecture notes, two volumes. Budapest, Tankönyvkiadó.
- , 1967, *On some philosophical problems of mathematics + Mathematical logic* (the latter with János Urbán). Budapest, Tankönyvkiadó.
- , 1968, *Between mathematics and philosophy*. Budapest, Gondolat. (German: *Die Begriffswelt der Mathematik*. 1976, Berlin, Volk und Wissen.)
- , 1969, *Elementary logic*. Lecture notes. Budapest, Tankönyvkiadó.
- , 1973, *Symbolic logic*. Lecture notes, three volumes (with co-authors). Budapest, Tankönyvkiadó.
- , 1973b, Prior-type modal logic (in English). *Periodica Mathematica Hungarica*. First part: 51-69., second part: 183-201.
- , 1979, A hundred years of symbolic logic: the oeuvre of Gottlob Frege. *Magyar Filozófiai Szemle*. 590-613.
- , 1980, Intensional logic without intensional variables. In I. Ruzsa (ed.), *Modal and intensional logic*. Budapest, OM Marxizmus-Leninizmus Oktatási Főosztálya.
- , 1981, *Modal logic with descriptions* (in English). The Hague, M. Nijhoff.
- , 1984, *Classical, modal and intensional logic*. Budapest, Akadémiai Kiadó.
- , 1988, *Logical syntax and semantics I*. Budapest, Akadémiai Kiadó.
- , 1989, *Logical syntax and semantics II*. Budapest, Akadémiai Kiadó.
- , 1991, *Intensional logic revisited* (in English). Budapest, published by the author.
- , 1997, *Introduction to metalogic* (in English). Budapest, Áron Publishers.
- , 1998, *Introduction to modern logic* (with András Máté). Budapest, Osiris.

## Whose Logic is Three-Valued Logic?

**Abstract.** ‘It would be unfair to judge that I use a three-valued logic or that I abandon the principle of *tertium non datur*’, writes Imre Ruzsa in his (Ruzsa 1991, 11.). For Ruzsa a truth value gap (the “third” truth value) arises only from the “defects” of our expressions (for example when a definite description does not denote anything) and not because there are “gaps” in reality. In the first part of the paper we explain in some detail how the truth value gaps arise and how they are transmitted in Ruzsa’s system. In the second part we will argue that there may be sentences which in a sense reflects real gaps, in other words, that the third truth value is a real truth value.

### 1 THE SEMANTICS OF SEMANTIC VALUE GAPS

One of Imre Ruzsa’s main achievements in logic is his system of intensional logic with semantic value gaps. A semantic value gap arises when a well formed expression of our (natural or artificial) language fails to denote anything. The simplest case is perhaps a definite description without a denotation (e.g. ‘the present king of France’). In Ruzsa’s system there *are* denotations even in such cases—these are the artificial entities “filling” the gaps. The individual denoted by ‘the present king of France’ is not a real individual: Ruzsa’s choice is the set  $U$ , the set of “real” (actual and possible) individuals, simply because evidently  $U \notin U$ . A bit more precisely: the type  $\iota$  of individuals has the domain  $D(\iota) = U \cup \{U\}$ , and  $\Theta(\iota) = U$  is the type’s *zero entity*—the “object” denoted by e.g. the empty descriptions.

Systems of intensional logics have in general two kinds of semantic values: the extensions—in Ruzsa’s terminology, factual values—and the intensions. In the type  $o$  of sentences the factual values are the truth values, the intensions are functions from worlds and times (technically, from the set  $I = W \times T$ , where  $W$  is the set of possible worlds and  $T$  is the linearly ordered set of time moments) to

the factual values. The truth values are represented by  $3 = \{0, 1, 2\}$ ,  $\Theta(o) = 2$  being the zero entity representing the truth value gap. If a sentence  $p$  has 2 as factual value (in a world  $w$  in a given moment  $t$ ) than we say it has no “real” truth value (in the world  $w$  in the moment  $t$ ). The “real” truth values are of course 0 (representing falsity) and 1 (representing truth).

From the basic types  $\iota$  and  $o$ , we get the other (functor) types. For example, predicates are expressions of type  $o(\iota)$ . The domain of this type is the set of all functions  $f : D(o) \rightarrow D(\iota)$  for which  $f(U) = 2$ . The zero entity of this type is the constant function having 2 for all the arguments. A predicate is partial if its interpretation  $\sigma(P) : D(\iota) \rightarrow 3$  maps more than one individual to 2.

In what follows  $|A|_{vi}$  denotes the factual value of  $A$  at the *index*  $i = \langle w, t \rangle \in W \times T$  according to the valuation  $v$ . If  $i$  is an index,  $d(i) \subseteq D(\iota)$  is the set of the “actual” individuals at  $i$ , that is, the actual individuals in the world  $w$  at the time moment  $t$ . If  $x$  is a variable of type  $\iota$ , than the value  $v(x)$  is always an element of  $D(\iota)$ ; if  $v(x) \notin d(i)$ , then  $|x|_{vi} = \Theta(\iota)$  and similarly for constants of type  $\iota$ . It could happen that a value of variable in the world  $w$  at the moment  $t$  is an individual not belonging to the domain of  $w$  at  $t$ —in such cases the factual value of the variable is the zero entity of the type.

Definite descriptions are handled as it is expected. If  $F$  is of type  $o(\iota)$  then the factual value of  $\mathbb{I}F$  (‘the  $F$ ’) is  $|\mathbb{I}F|_{vi} = u_i$  if  $\{u \in d(i) : |F|_{vi}(u) = 1\} = \{u_i\}$ , and in all other cases  $|\mathbb{I}F|_{vi} = \Theta(\iota)$ . If there is exactly one  $F$  in the world  $w$  at the moment  $t$ , then  $|\mathbb{I}F|_{vi}$  is *this* object; and if the set of the  $F$ s is empty or has more than one element,  $|\mathbb{I}F|_{vi}$  is the zero entity.

The identities are (of course) expressions of type  $o$ . If  $A$  and  $B$  are of the same type  $\alpha$ , then

$$|A = B|_{vi} = \begin{cases} 2 & \text{if } |A|_{vi} = \Theta(\alpha) \text{ or } |B|_{vi} = \Theta(\alpha) \\ 1 & \text{if } |A|_{vi} = |B|_{vi} \neq \Theta(\alpha) \\ 0 & \text{otherwise} \end{cases}$$

According to this rule, if on one side of an identity stands an expression having the zero entity of its type as its factual value then the identity’s factual value will be automatically 2. It has the (somewhat strange) consequence that non-existent individuals cannot be identical even with themselves. For example the sentence ‘the present king of France’ = ‘the present king of France’ falls in the truth value gap.

Ruzsa—following an idea of Tarski—defines the propositional connectives in terms of  $\lambda$  and  $=$ . The technical details (see (Ruzsa 1991, 39–41)) do not concern us, only the truth tables governing the connectives. The truth tables for negation, conjunction, and alternation are the following:

$p$	$\sim p$
1	0
0	1
2	2

$p \wedge q$	1	0	2
1	1	0	2
0	0	0	2
2	2	2	2

$p \vee q$	1	0	2
1	1	1	2
0	1	0	2
2	2	2	2

The tables for the conditional and the biconditional (the latter is simply the = in the type  $o$ ):

$p \supset q$	1	0	2
1	1	0	2
0	1	1	2
2	2	2	2

$p \equiv q$	1	0	2
1	1	0	2
0	0	1	2
2	2	2	2

These tables are the weak Kleene tables. The connectives working according to them always transmit the truth value gap from the part to the whole. This is not true for the strong Kleene connectives for which the truth tables are the following:

$p$	$\neg p$
1	0
0	1
2	2

$p \& q$	1	0	2
1	1	0	2
0	0	0	0
2	2	0	2

$p \vee q$	1	0	2
1	1	1	1
0	1	0	2
2	1	2	2

$p \supset q$	1	0	2
1	1	0	2
0	1	1	1
2	1	2	2

$p \equiv q$	1	0	2
1	1	0	2
0	0	1	2
2	2	2	2

The strong Kleene conjunction, alternation, and conditional does not transmit the truth value gap: for example, if  $p$  is true then the value of  $p \vee q$  is 1 (true) even if the truth value of  $q$  is 2. The strong Kleene connectives are in better harmony with “the logic of empirical investigations”, the conception that Ruzsa calls *epistemic*. In such a logic 2 denotes the value “yet unknown”.

Why did Ruzsa decide in favor of the weak versions? The question has (at least) three answers. One is a bit personal: the epistemic conception reminds him of the “so called” intuitionist logic; and this logic, according to his conception, does not even deserve the name ‘logic’.<sup>1</sup> The second, and less personal, answer is that using Ruzsa’s system’s temporal operators and introducing an epistemic operator, one could probably succeed in modeling some aspects of the “epistemic conception”.

<sup>1</sup>He once told us the following story. In a conference (probably in the sixties) when he used the arrow for the conditional, the chair asked him: “So you are an intuitionist, aren’t you?” At the very moment he decided to use the horseshoe symbol: let there be no mistake.

But Ruzsa's most prominent reason is the importance of the transmittal of the semantic value gaps, a phenomenon we have already seen in the definitions of the domain of functor types and that of the factual values of the identites. A general theorem of his system states that this phenomenon holds in general.<sup>2</sup>

Summing up: in extensional contexts, a semantic value gap is a special "illness" for which the treatment is: not allowing it to disappear. A truth value gap is really a *gap*, arising from clashes of language and reality; it is impossible for the "real world" to have gaps.

My *credo* is simply this: A sentence may or may not have a truth value. If it has one then it expresses a statement which is either true or false. The lack of a truth value is not a third truth value. (Ruzsa 1991, 11.)

## 2 ABSOLUTELY UNDECIDABLE SENTENCES

In the epistemic conception the semantic value gaps are due to the gaps in our logic. By contrast, Ruzsa's approach is ontological. As he puts it:

In (. . .) informal reasonings, the sources of value gaps are located in the realm of facts, in the formal semantics they [are] located in the interpretations of the (formal) language. (Ruzsa 1991, 12.)

The facts of which Ruzsa speaks are facts of the world *and* facts of our language, gaps only arise when something is mistaken in our expressions. Are there sources of truth value gaps "in the world" in which our language plays no significant role? In other words: are there gaps in reality? Before trying to answer this question, let's go back once more to the epistemic conception. According to our knowledge of it, every (well-formed, unambiguous) sentence  $p$  must fall in one of the following seven cases:

- (1)  $p$  is true, and we know (proved, verified) that it is true
- (2)  $p$  is true; we do not know that yet, but we will
- (3)  $p$  is true, but we will never know that it is true
- (4)  $p$  is absolutely undecidable (even God cannot determine its truth value)
- (5)  $p$  is false, but we will never know that it is false
- (6)  $p$  is false; we do not know that yet, but we will
- (7)  $p$  is false, and we know that it is false

In cases (1), (2), (6) and (7) there are no difficulties. Moreover, we have good candidates that are of case (3) or of case (5). For example, let  $p$  be the sentence

The value of the digit in the  $10^{10^{10}}$ th place of the decimal expansion of  $\pi - 3$  equal to zero.<sup>3</sup>

<sup>2</sup>If  $A$  (of type  $\alpha$ ) is an extensional component of  $B$  (of type  $\beta$ ), then  $|A|_{v_i} = \Theta(\alpha) \Rightarrow |B|_{v_i} = \Theta(\beta)$ , for the details see (Ruzsa 1991, 33.)

<sup>3</sup>Cf. (Feferman 2006).

The truth value of this sentence can be determined in principle by a mechanical check—but this check is far beyond our computational powers. Nevertheless, we can say that this sentence has a determinate truth value and God knows what it is.

What about case (4)? Does the existence of absolutely undecidable sentences threaten God's omniscience? We can say with Michael Dummett: not at all. God knows the answer to every question that has answer, and He knows of every question whether it has an answer. If there really are questions which has no answer even for Him then the divine logic must be three-valued, and instead of a truth value gap, there will be a genuine third truth value.<sup>4</sup>

In comparison with God, in this respect (too) we are in a more uncomfortable position: we cannot in principle distinguish cases (4), (5), and (6). The reason lies in what can be called the "Hauptsatz" of undecidable propositions: if  $p$  is an absolutely undecidable statement, then we cannot prove "constructively" that it is really the case. The proof (due to Martin-Löf<sup>5</sup>) relies heavily on the constructivist (verificationist) conception of negation: proving  $\neg p$  amounts to showing that any attempt to prove  $p$  will eventually be blocked (in mathematics, by a contradiction; in general, by some serious difficulty). Proving that  $p$  is undecidable amounts to a proof that any attempt to prove  $p$ , as well as any attempt to prove  $\neg p$ , will eventually be blocked. But it is nothing but a proof of  $\neg p$ , and  $\neg\neg p$ , respectively. We arrive at a contradiction (a serious difficulty).

So we have to rely on our intuitions.

### 3 FINITE KNOWLEDGE OF THE INFINITE

We cannot prove of a sentence that it is absolutely undecidable, but we can perhaps imagine what such a sentence could be. Feferman's (in fact, the intuitionists') example about the decimal expansion of  $\pi$  gives a clue. If we—with our limited means—cannot end a process that is too long, it is conceivable that an infinite process is such that even an infinite mind cannot go through it.

But we must be careful. Even we, finite beings know very much about the natural numbers, we can prove for example that for every natural number  $n$ ,  $7^n$  is divisible with 6, that there are infinitely many primes and so on. We have methods which make the infinite finite, that is, methods (first of all, mathematical induction) by which we can prove, say, that for some property  $F$ ,

<sup>4</sup>See e.g. (Dummett 2006, 108—109).

<sup>5</sup>(Martin-Löf 1995). Martin-Löf actually "proves" that there are no undecidable propositions. His proof relies on the intuitionist conceptions of proposition, truth, falsity, and knowledge. For someone not in the intuitionist camp, his conclusion can be formulated as follows: there may be absolutely undecidable propositions, but we cannot produce one about which we can prove that it is really absolutely undecidable. Cf. (Feferman 2006, 147.).

there are infinitely many numbers for which it holds. Augustine and many of his followers believe that for God it is so with *every* property  $F$ .

As for their other assertion, that God's knowledge cannot comprehend things infinite, it only remains for them to affirm, in order that they may sound the depths of their impiety, that God does not know all numbers. For it is very certain that they are infinite. . . Does God, therefore, not know numbers on account of this infinity; and does His knowledge extend only to a certain height in numbers, while of the rest He is ignorant? Who is so left to himself as to say so?

. . . if everything which is comprehended is defined or made finite by the comprehension of him who knows it, then all infinity is in some ineffable way made finite to God, for it is comprehensible by His knowledge. (Augustine 1993, Book XII., Chapter 18.)

According to this conception, the property ' $n$  is one member of a twin prime-pair' is for God as simple as for us the property ' $n$  is a prime number': He can decide whether it holds for infinitely many numbers or not. He can perhaps "see" some higher-order structure which decide the matter (as in the case of Fermat's last theorem there are structures revealed by the theory of analytic functions that decide that a property holds for all natural numbers bigger than 2). But is it really the case? Is every property of natural numbers such that it is in principle possible "making it finite"—deciding in finite steps, whether it holds for infinitely many numbers or not? If there is a property  $F$  for which even God has no other choice in order to decide whether it holds for infinitely many numbers than to check "all" the numbers one after another, then the sentence 'there are infinitely many numbers  $n$  for which  $F(n)$  holds' is a good candidate for being absolutely undecidable.

We can argue that—*pace* Augustine—in this case even God cannot determine the truth value of this sentence (but He would know *that*). Such a sentence would then be per definitionem absolutely undecidable, and as such, it would signal the presence of a real gap "in the world".

#### 4 INFINITE TASKS

What makes it impossible even for God to run through an infinite series of computations? The strongest argument can be extracted from the paradoxes of super-tasks. The classic example of these paradoxes is Thomson's lamp.

There are certain reading-lamps that have a button in the base. If the lamp is off and you press the button the lamp goes on, and if the lamp is on and you press the button the lamp goes off.

Suppose now that the lamp is off, and I succeed in pressing the button an infinite number of times, perhaps making one jab in one minute, another



jab in the next half-minute, and so on. . . After I have completed the whole infinite sequence of jabs, i.e., at the end of two minutes, is the lamp on or off? It seems impossible to answer this question. It cannot be on, because I did not ever turn it on without at once turning it off. It cannot be off, because I did in the first place turn it on, and thereafter I never turned it off without at once turning it on. But the lamp must be either on or off. This is a contradiction.<sup>6</sup>

One lesson from the paradox is simply this: carrying out a super-task is *conceptually* impossible. For “at the end” of the infinite series of tasks, there could be a discontinuity which cannot be explained. Russell famously called it only “medically impossible” running through the whole expansion of  $\pi$ , see (Russell 1953, 143.). By contrast, Michael Dummett argues that “the reason why we cannot survey an infinite totality is not the deficiency of human capabilities: it is that it is *senseless* to imagine an infinite task completed” (Dummett 2006, 70–71; the italic is Dummett’s).

Arithmetic can be a natural realm of super-tasks. If there is a property  $F$  of natural numbers for which the truth value of  $F(n)$  for each number  $n$  can be determined by finite computation but there are “absolutely” no general method for determining whether  $F(n)$  holds for infinitely many numbers or not then even a “Divine Arithmetician” who can carry out a computation “infinitely quickly” cannot determine the truth value of the statement *there are infinitely many  $n$  for which  $F$  holds* or, of the statement *there are infinitely many  $n$  for which  $F$  does not hold*. For to decide these statements, She has to check every number  $n$ , that is, by running through an infinity of tasks. And this is impossible—even for the Divine Arithmetician. The reason is that the same kind “discontinuity” would arise as in the case of the Thomson’s lamp. For suppose the “prover” has a white paper. After checking  $F(0)$ , she paint it black; then after checking  $F(1)$  she paint it again white; and so on. (If she can decide whether  $F(n)$  is true or false then manipulating the paper is only a simple extra.) What color will be the paper after checking all of the natural numbers? There is no answer—the super-task cannot be carried out.<sup>7</sup>

<sup>6</sup>Thomson (1954), cited in (Sainsbury 2009, 12.). The paradox resembles that of the staccato run, a variant of Zeno’s Racetrack paradox. In the staccato version the runner - say, Achilles - runs for half a minute, then pauses for half a minute, then runs for a quarter of a minute, then pauses for a quarter of a minute, and so on ad infinitum. At the end of two minutes he will have stopped and started in this way infinitely many times. Each time he pauses he could perform a task of some kind. Then at the end of two minutes he will have performed infinitely many of these tasks. On the staccato run and other paradoxes of the infinite, see (Moore 1990).

<sup>7</sup>It is a (super-)task for the philosophers of time to explain exactly what makes it impossible for (even) the Divine Arithmetician to run through an infinite series of tasks. For if the continuum of time has the structure—say—of the real numbers than for “someone” who could count with no speed limit, it may be possible to determine whether  $F(0)$  holds or not in a minute,  $F(1)$  in half a minute,  $F(2)$  in a quarter of a minute and so on. . .

If  $p$  is such a sentence then it may happen that God does not know whether it is true or false. (But even in this case He knows *that*.) And in this case we can call 2 a *real* truth value.

## 5 WHAT WOULD RUZSA SAY

Without any doubt this argument would not affect Ruzsa's philosophical position. If there was a knock-down argument against his—Platonist—view from the standpoint of the constructivists and the intuitionists, it would not be a philosophical argument like the preceding one. And I can imagine Imre Ruzsa stamping his foot saying: “after all, there are infinitely many  $n$  for which  $F(n)$  is true or there are only finitely many such  $n$ , there is no third possibility”.\*

## REFERENCES

- Augustine (Aurelius Augustinus), 1993, *City of God*. Transl. Marcus Dods. New York, The Modern Library.
- Dummett, M., 2006, *Thought and Reality*. Oxford, Clarendon Press.
- Feferman, S., 2006, Are there absolutely unsolvable problems? *Philosophia Mathematica* 14 (III), 134–152.
- Martin-Löf, P., 1995, Verificationism then and now. In DePauli-Schimanovich, W. & al. (eds.), *The Foundational Debate*. Dordrecht, Kluwer, 187–196.
- Moore, A. W., 1990, *The Infinite*. Oxford, Routledge.
- Russell, B., 1953, The Limits of Empiricism. *Proceedings of the Aristotelian Society* 36.
- Ruzsa, I., 1991, *Intensional Logic Revisited*. Budapest, published by the author.
- Sainsbury, R. M., 2009, *Paradoxes*. 3rd ed. Cambridge, Cambridge University Press.
- Thomson, J. F., 1954, Tasks and super-tasks. *Analysis* 15, 1–13.

\*The author happily acknowledges the support of the Philosophy of Language Research Group of the Hungarian Academy of Sciences.

# On Models of General Type-Theoretical Languages

**Abstract.** In the present paper we consider general type theoretical languages as the representations of the functor–argument decomposition and compositional semantics relying on it and find some theorems making explicit the theoretical presuppositions of general type theoretical languages and their total or partial semantics. After defining the notion of semantic categories in the spirit of Husserl, we characterize Tarskian and Husserlian models both in total and partial semantics and prove their characteristic theorems.

## 1 PERSONAL FOREWORD

I am greatly indebted to Professor Imre Ruzsa for the opportunity to work with him for almost two decades. After graduation I began to work as a research assistant at Kossuth University, Debrecen in 1979 and I wrote a letter to professor Imre Ruzsa. In spite of the fact that we had never met and did not know each other personally, he answered soon. The first personal meeting changed my scientific life profoundly. I have no opportunity to tell the whole story, but I should like to emphasize that I should be quoting his books and papers<sup>1</sup> in almost each sentence of the present paper, which is dedicated to the memory of Professor Imre Ruzsa.

## 2 INTRODUCTION

From the theoretical point of view, type theoretical languages (with a lambda operator) represent function abstraction and function application and rely on functor–argument decomposition, which goes back to Frege.

<sup>1</sup>I mention here only three of them: (Ruzsa 1989), (Ruzsa 1991), (Ruzsa 1997).

In Frege's view, one of the most important inventions of *Begriffsschrift* is the replacement of the subject–predicate decomposition by the functor–argument one. He wrote the following: “The very invention of this *Begriffsschrift*, it seems to me, has advanced logic. . . . [L]ogic hitherto has always followed ordinary language and grammar too closely. In particular, I believe that the replacement of the concept *subject* and *predicate* by *argument* and *function* will prove itself in the long run. It is easy to see how taking a content as a function of an argument gives rise to concept formation. . . . The distinction between subject and predicate finds no place in my representation of a judgement.”<sup>2</sup> (Frege 1879/1997, 51, 53.)

One of the most general theoretical representations of the functor–argument decomposition is the well-known type theory (or the different systems of type–theoretical language and/or logic<sup>3</sup>).

Generally, syntactic categories have to be distinguished from semantic ones. At the same time, our formal systems fulfill the following fundamental principle of formal type–theoretical semantics:

[The mirror principle:] Associated with every syntactic category  $C$  is a counterpart semantic category  $C^*$ , whose *mathematical type* ‘mirrors’ the *grammatical type* of  $C$ . And, in particular, every expression of syntactic category  $C$  is interpreted by an object of semantic category  $C^*$ .  
(Dunn and Hardegree 2001, 142.)

On the basis of the mirror principle, in what follows, we are speaking about types, and using them to define and denote different syntactic categories and the corresponding sets of possible semantic values.

### 3 GENERAL FORMAL SYSTEM

At first, the system of types has to be defined. The system of types relies on primitive type(s). Generally we have only one requirement: the symbol  $o$  must be a primitive type. From the theoretical point of view, the main reason for this is that the symbol  $o$  is taken as the type of the most fundamental expressions of our formal language. Expressions of type  $o$  are called formulae. Formulae directly correspond to a special sort of conceptual content or information. It means that formulae are the structures of complete information or closed (and whole) conceptual content. In a given interpretation, formulae are intended to represent complete information called proposition in the literature.

There is another, mainly semantic reason for type  $o$  having been declared to be primitive. From the semantic point of view, Frege's context principle or as

<sup>2</sup>I use the expression ‘functor’ instead of ‘function’ in order to differentiate an incomplete expression of a language from its semantic value.

<sup>3</sup>It goes back to (Church 1940).

(Hodges 2001a) says, Frege's Dictum can be taken as a general leading idea. In *The Foundations of Arithmetic* Frege wrote the following, usually quoted as the context principle:

never to ask for the meaning of a word in isolation, but only in the context of a proposition; (Frege 1884/1980, x.)

It is enough if the proposition taken as a whole has sense; it is this that confers on its parts also their content. (Frege 1884/1980, 71.)

According to the context principle, an expression has sense (meaning) only in the sentence in which it occurs. Sometimes we need more than one primitive type (usually individual names constitute another primitive type). The main difference between primitive and non-primitive types is that the semantic domains of primitive types have to be given via definition, while the domains of non-primitive types are originated from them. Non-primitive types are usually called functor types.

**Definition 1.** Let  $PT$  be an arbitrary set of symbols, the set of primitive types, such that  $o \in PT$ . Then the set  $TYPE_{PT}$  is defined inductively as follows:

- (1)  $PT \subseteq TYPE_{PT}$ ;
- (2)  $\alpha, \beta \in TYPE_{PT} \Rightarrow \langle \alpha, \beta \rangle \in TYPE_{PT}$ .

**Remark 1.** Here  $o$  is the type of formulae from the syntactic point of view, and the type of their possible semantic values from the semantic point of view.  $\langle \alpha, \beta \rangle$  is the type of functors which, when they are filled in by an argument of type  $\alpha$ , yield an expression of type  $\beta$  in syntax (in the formal language), and it stands for the type of function from objects of type  $\alpha$  to objects of type  $\beta$  in semantics.

The type-theoretical language is the most general one concerning the functor-argument decomposition. It has only two syntactic operations: filling a functor with an argument (function application from the semantic point of view) and lambda abstraction. The latter produces a way to create a functor from an expression.

**Definition 2.** A type-theoretical language is an ordered quadruple

$$L = \langle LC, Var, Con, Cat \rangle$$

satisfying the following conditions:

- (1)  $LC$  is the set of theoretical constants.<sup>4</sup>  $LC = \{\lambda, (, )\}$
- (2)  $Var = \bigcup_{\alpha \in TYPE_{PT}} Var(\alpha)$  and  $Var(\alpha)$  is a denumerably infinite set of symbols<sup>5</sup>.

<sup>4</sup>A theoretical constant has the same semantic value (or sense) in every interpretation as a logical constant does in a logical system.

<sup>5</sup> $Var(\alpha)$  is the set of variables of the type  $\alpha$ .

- (3)  $Con = \bigcup_{\alpha \in TYPE_{PT}} Con(\alpha)$ , where  $Con(\alpha)$  is a denumerably set of symbols.<sup>6</sup>  
 (4) All mentioned sets of symbols are assumed to be pairwise disjoint ones.  
 (5)  $Cat = \bigcup_{\alpha \in TYPE_{PT}} Cat(\alpha)$ , where the sets  $Cat(\alpha)$  are defined by the inductive rules (a) . . . (c) as follows:<sup>7</sup>  
 (a)  $Var(\alpha) \cup Con(\alpha) \subseteq Cat(\alpha)$ ;  
 (b)  $C \in Cat(\langle \alpha, \beta \rangle)$ ,  $B \in Cat(\alpha) \Rightarrow 'C(B)' \in Cat(\beta)$ ;  
 (c)  $A \in Cat(\beta)$ ,  $\tau \in Var(\alpha) \Rightarrow '(\lambda \tau A)' \in Cat(\langle \alpha, \beta \rangle)$ ;

The (total or partial) functor–argument frame is the compositional mirror of a type–theoretical language. It can be said that the functor–argument frame gives possible semantic values.

**Definition 3.** A total functor–argument frame  $F$  is the system of sets  $\langle Dom_F(\gamma) \rangle_{\gamma \in TYPE_{PT}}$  such that

- (1) If  $\gamma \in PT$ , then  $Dom_F(\gamma)$  is an arbitrary nonempty set.  
 (2)  $Dom_F(\langle \alpha, \beta \rangle) = Dom_F(\beta)^{Dom_F(\alpha)}$  for all  $\langle \alpha, \beta \rangle \in TYPE_{PT}$

**Definition 4.** A partial functor–argument frame  $PF$  is the system of sets  $\langle Dom_{PF}(\gamma) \rangle_{\gamma \in TYPE_{PT}}$  such that

- (1) if  $\gamma \in PT$ , then  $Dom_{PF}(\gamma)$  is an arbitrary set with a distinguished member  $\Theta_\gamma$ , which is called the null entity of type  $\gamma$ , such that  $Dom_{PF}(\gamma) \setminus \{\Theta_\gamma\} \neq \emptyset$ ;  
 (2)  $Dom_{PF}(\langle \alpha, \beta \rangle) = Dom_{PF}(\beta)^{Dom_{PF}(\alpha)}$  for all  $\langle \alpha, \beta \rangle \in TYPE_{PT}$  and  $\Theta_{\langle \alpha, \beta \rangle} = g$  where  $g \in Dom_{PF}(\langle \alpha, \beta \rangle)$  and  $g(u) = \Theta_\beta$  for all  $u \in Dom_{PF}(\alpha)$ .

Interpretive function and assignment associate the constants and the variables of the type–theoretical language with their semantic values. In a model, which consists of a frame, an interpretive function and an assignment, semantic rules can be defined to determine the semantic values of compound expressions with respect to the given model.

**Definition 5.** A (total or partial) model  $M$  on  $G$  is an ordered triple  $\langle G, \varrho, v \rangle$  where

- (1)  $G$  is a (total or partial) functor–argument frame;  
 (2)  $\varrho, v$  are functions with domains  $Con$  and  $Var$  respectively<sup>8</sup> such that  
 (a) if  $a \in Con(\alpha)$ , then  $\varrho(a) \in Dom_G(\alpha)$ ;  
 (b) if  $\tau \in Var(\alpha)$ , then  $v(\tau) \in Dom_G(\alpha)$ .

**Remark 2.** A model  $M$  on  $G$  is total or partial if  $G$  is a total or partial functor–argument frame respectively.

<sup>6</sup> $Con$  is the set of non–theoretical symbols of  $L$ . The semantic value of an expression belonging to the set  $Con$  is given by an interpretation.

<sup>7</sup> $Cat$  is the set of all well–formed expressions of  $L$ . The set  $Cat(\alpha)$  is the  $\alpha$ –category of  $L$  ( $\alpha \in TYPE_{PT}$ ).

<sup>8</sup> $\varrho$  is an interpretive function,  $v$  is an assignment.

If  $M = \langle F, \varrho, v \rangle$  is a total model on  $F$ , then

$$\text{Dom}_M(\alpha) = \text{Dom}_F(\alpha).$$

If  $PM = \langle PF, \varrho, v \rangle$  is a partial model on  $PF$ , then

$$\text{Dom}_{PM}(\alpha) = \text{Dom}_{PF}(\alpha) \setminus \{\Theta_\alpha\}.$$

If  $M = \langle G, \varrho, v \rangle$  is a total or partial model,  $\xi \in \text{Var}(\gamma)$  and  $u \in \text{Dom}_G(\gamma)$ , then the model  $M_\xi^u = \langle G, \varrho, v[\xi : u] \rangle$  is like  $M$  except that  $v[\xi : u](\xi) = u$ .

**Definition 6.** A total or partial model  $M = \langle G, \varrho, v \rangle$  assigns each expression  $A$  of type  $\alpha$  a semantic value  $\llbracket A \rrbracket_M$  according to the following semantic rules:

- (1) if  $a \in \text{Con}(\gamma)$ , then  $\llbracket a \rrbracket_M = \varrho(a)$ ;
- (2) if  $\xi \in \text{Var}(\gamma)$ , then  $\llbracket \xi \rrbracket_M = v(\xi)$ ;
- (3) if  $A \in \text{Cat}(\langle \alpha, \beta \rangle)$  and  $B \in \text{Cat}(\alpha)$ , then  $\llbracket A(B) \rrbracket_M = \llbracket A \rrbracket_M(\llbracket B \rrbracket_M)$ ;
- (4) if  $A$  is an expression of type  $\beta$  and  $\xi \in \text{Var}(\alpha)$ , then  $\llbracket \lambda \xi A \rrbracket_M = g$ , where  $g$  is a function from  $\text{Dom}_G(\alpha)$  to  $\text{Dom}_G(\beta)$  such that  $g(u) = \llbracket A \rrbracket_{M_\xi^u}$  for all  $u \in \text{Dom}_G(\alpha)$ .

**Proposition 1.** If  $M$  is a total model and  $A \in \text{Cat}(\alpha)$ , then  $\llbracket A \rrbracket_M \in \text{Dom}_M(\alpha)$ . If  $M$  is a partial model, then  $\llbracket A \rrbracket_M \in \text{Dom}_M(\alpha) \cup \{\Theta_\alpha\}$ .

**Definition 7.** If  $M$  is a total or partial model, then  $A$  is meaningful with respect to  $M$ , in symbols  $A \in \text{Cat}_{mf}^M$  if  $A \in \text{Cat}(\alpha)$  for some type  $\alpha$  and  $\llbracket A \rrbracket_M \in \text{Dom}_M(\alpha)$ .

**Remark 3.** If  $M$  is a total model, then all  $A \in \text{Cat}$  are meaningful, i.e. there is no difference at all between the notions of well-formedness and meaningfulness. We can only make a real differentiation between them in the case of partial models.

**Theorem 1.** If  $A \in \text{Cat}$ ,  $M_1 = \langle G, \varrho, v_1 \rangle$  and  $M_2 = \langle G, \varrho, v_2 \rangle$  are two (total or partial) models of  $L$  with the same frame  $G$  and interpretive function  $\varrho$  such that  $v_1(\tau) = v_2(\tau)$  for all  $\tau \in V(A)^9$ , then  $\llbracket A \rrbracket_{M_1} = \llbracket A \rrbracket_{M_2}$ .

**Proposition 2.** If  $A \in \text{Cat}$  is a closed expression, then  $\llbracket A \rrbracket_M$  is independent from  $v$  i.e.  $\llbracket A \rrbracket_M = \llbracket A \rrbracket_{M_\tau^u}$  for all  $\tau \in \text{Var}(\gamma)$  and  $u \in \text{Dom}_F(\gamma)$ .<sup>10</sup>

To prove lambda-conversion law, we need the Law of replacement 2 and Lemma 1. The first one says that in semantics, we only take into consideration semantic values and don't pay any attention to the expression itself—except its type—whose semantic value is given. It doesn't matter how we get a

<sup>9</sup>The definitions of subterms, free variables, open and close expressions and the substitutability are usual ones. The set  $V(A)$ , is the set of free variables of the expression  $A$ .

<sup>10</sup>In the case of closed expressions we can speak about models as ordered pairs of frames and interpretive functions.

semantic value, what form of the compound expression gets the semantic value. We may formulate the property in the law of replacement by means of universal replacement of expressions belonging to the same type with the same semantic value. >From the logical–philosophical point of view, the law of replacement is a special type–theoretical formulation of a version of the principle of compositionality called the substitutivity principle, which goes back to Leibniz.

[The Substitutivity Principle:] If two expressions have the same meaning, then substitution of one for the other in a third expression does not change the meaning of the third expression. (Szabó 2000, 490.)

I must emphasize that the law of replacement can only be considered as a restricted version of the substitutivity principle, the unrestricted form of the substitutivity principle holds only in Husserlian models dealt with in Section 6. The next definition introduces the notion of 1–compositionality. 1–compositional systems fulfill a restricted version of the substitutivity principle, and Corollary 1 of Law of replacement 2 says that our general system is compositional in the sense of 1–compositionality.

**Definition 8.** *Let  $M$  be a model of  $L$ . We say that  $M$  is 1–compositional if for all well–formed expressions  $A, B, C$  ( $A, B, C \in \text{Cat}$ ) and variable  $\tau$  ( $\tau \in \text{Var}$ ) such that  $(\lambda\tau C)(A), (\lambda\tau C)(B) \in \text{Cat}_{mf}^M$*

$$\llbracket A \rrbracket_M = \llbracket B \rrbracket_M \Rightarrow \llbracket (\lambda\tau C)(A) \rrbracket_M = \llbracket (\lambda\tau C)(B) \rrbracket_M$$

**Theorem 2** (Law of replacement).<sup>11</sup>

*If  $A \in \text{Cat}$ ,  $B, C \in \text{Cat}(\gamma)$ , and  $M$  is a (total or partial) model of  $L$ , then*

$$\llbracket B \rrbracket_M = \llbracket C \rrbracket_M \Rightarrow \llbracket A \rrbracket_M = \llbracket A[C \downarrow B] \rrbracket_M.$$

**Corollary 1.** *If  $M$  is a (total or partial) model of  $L$ , then  $M$  is 1–compositional.*

**Lemma 1.** *If  $B$  is substitutable for variable  $\tau$  in  $A$ ,  $M$  is a (total or partial) model, and  $\llbracket B \rrbracket_M = u$ , then  $\llbracket A_\tau^B \rrbracket_M = \llbracket A \rrbracket_M^u$ .*

**Theorem 3** (Lambda–conversion law). *If  $A \in \text{Cat}$ ,  $\tau \in \text{Var}(\beta)$ ,  $B \in \text{Cat}(\beta)$  and  $B$  is substitutable for  $\tau$  in  $A$ , then  $\llbracket (\lambda\tau A)(B) \rrbracket_M = \llbracket A_\tau^B \rrbracket_M$  for all (total or partial) models  $M$ .*

#### 4 PROPERTIES OF TOTAL AND PARTIAL MODELS

Let us turn our attention to different, total or partial models.<sup>12</sup> We need some notions to compare and combine different models. In the following

<sup>11</sup>If  $A \in \text{Cat}$  and  $B, C \in \text{Cat}(\gamma)$ , then  $A[C \downarrow B]$  ( $\in \text{Cat}$ ) is obtained by replacing a subterm occurrence (i.e. not preceded immediately by  $\lambda$ ) of  $B$  by  $C$ .

<sup>12</sup>The proofs of theorems in Section 4.5 can be found in (Mihálydeák 2010, 127–131.).



definitions let  $L (= \langle LC, Var, Con, Cat \rangle)$  be a type-theoretical language and  $M (= \langle G, \varrho, v \rangle)$  be its total or partial model.

**Definition 9.**

- (1) If  $\approx$  is an equivalence relation on the set  $Cat'$  ( $\subseteq Cat$ ), then  $\approx$  is a synonymy for  $L$ . The set  $Cat'$  is the field of synonymy  $\approx$ .
- (2) Syntactic synonymy for  $L$  is the synonymy  $\cong_L$  generated by the syntax of  $L$ , i.e.  $A \cong_L B$  if and only if there is a type  $\gamma$  such that  $A, B \in Cat(\gamma)$ .
- (3) Synonymy generated by the model  $M$  is a synonymy  $\approx_M$  for  $L$  with the field  $Cat_{mf}^M$  such that  $A \approx_M B \Leftrightarrow \llbracket A \rrbracket_M = \llbracket B \rrbracket_M$ .
- (4) Closed synonymy (or  $c$ -synonymy) generated by the model  $M$  is a synonymy  $\approx_{Mc}$  for  $L$  with the field  $\{A : A \in Cat, A \text{ is closed}\} \cap Cat_{mf}^M$  such that  $A \approx_{Mc} B \Leftrightarrow \llbracket A \rrbracket_M = \llbracket B \rrbracket_M$ .
- (5) A synonymy  $\approx$  for  $L$  is semantic if there is a model  $M$  of  $L$  such that  $\approx_M$  equals  $\approx$ .

The next proposition shows that in a general type-theoretical compositional framework, syntactic synonymy can be treated as a degenerate semantic one.

**Proposition 3.** *The syntactic synonymy for  $L$  is semantic (in a degenerate sense).*

**Remark 4.** *In what follows, a model of  $L$  generating the synonymy  $\cong_L$  is denoted by  $M_L$  and called 'syntactic' model.*

**Definition 10.**

- (1) Two models  $M_1, M_2$  of a language  $L$  are said to be equivalent (closed equivalent,  $c$ -equivalent) if  $\approx_{M_1}$  equals  $\approx_{M_2}$  ( $\approx_{M_1c}$  equals  $\approx_{M_2c}$ ), i.e. their generated synonymies ( $c$ -synonymies) are equivalent.
- (2) Given two synonymies  $\approx$  and  $\approx'$  for  $L$ , we say that  $\approx$  is compatible with  $\approx'$  if for all expressions  $A, B$  ( $\in Cat$ ) in the field of both synonymies,  $A \approx B \Leftrightarrow A \approx' B$ .
- (3) Given two synonymies  $\approx$  and  $\approx'$  for  $L$ , we say that  $\approx$  is closed compatible with (or  $c$ -compatible with)  $\approx'$  if for all closed expressions  $A, B$  ( $\in Cat$ ) in the field of both synonymies  $A \approx B \Leftrightarrow A \approx' B$ .
- (4) We say that two models  $M_1, M_2$  of  $L$  are compatible (closed compatible) if their generated synonymies  $\approx_{M_1}, \approx_{M_2}$  are compatible ( $c$ -compatible).

**Proposition 4.** *If the models  $M_1, M_2$  of  $L$  are equivalent, then  $M_1$  and  $M_2$  are compatible and  $c$ -compatible.*

**Proposition 5.** *If the models  $M_1, M_2$  of  $L$  are equivalent, then  $M_1$  and  $M_2$  are  $c$ -equivalent.*

**Proposition 6.** *If  $M (= \langle G, \varrho, v \rangle)$  is a model of  $L$ ,  $\tau \in Var(\gamma)$  and  $u \in Dom_G$ , then the models  $M$  and  $M_\tau^u$  are  $c$ -equivalent.*

**Proposition 7.** *If the models  $M_1, M_2$  of  $L$  are compatible, then  $M_1, M_2$  are  $c$ -compatible.*

**Proposition 8.** *Let the models  $M_1, M_2$  of  $L$  be total.  $M_1, M_2$  are compatible if and only if  $M_1, M_2$  are equivalent.*

In order to investigate the connection between total and partial semantic systems, we need a ‘total’ or ‘pseudo partial’ part of a partial frame  $PF$ , which is denoted by  $PF^t$ .

**Definition 11.** *Let  $PF$  be a partial frame. The total part  $PF^t$  of the partial frame  $PF$  is the system of sets  $\langle Dom_{PF}^t(\gamma) \rangle_{\gamma \in TYP E_{PT}}$  such that*

- (1) *if  $\gamma \in PT$ , then  $Dom_{PF}^t(\gamma) = Dom_{PF}(\gamma) \setminus \{\Theta_\gamma\}$ ;*
- (2) *if  $\gamma = \langle \alpha, \beta \rangle$  then  $Dom_{PF}^t(\gamma) \subseteq Dom_{PF}(\gamma)$  such that for all  $f \in Dom_{PF}^t(\langle \alpha, \beta \rangle)$   $f(u) \in Dom_{PF}^t(\beta)$  if  $u \in Dom_{PF}^t(\alpha)$  and  $f(u) = \Theta_\beta$  otherwise.*

**Remark 5.**

- (1) *For the sake of brevity, we use the notation ‘ $Dom_F^t$ ’ in the case of a total frame  $F$ . Of course, in this case  $Dom_F^t(\gamma) = Dom_F(\gamma)$  for all  $\gamma \in TYP E_{PT}$ .*
- (2) *If  $M (= \langle G, \varrho, \nu \rangle)$  is a total or partial model, then  $Dom_M^t(\gamma) = Dom_G^t(\gamma)$  for all  $\gamma \in TYP E_{PT}$ .*

**Definition 12.** *An expression  $A$  of type  $\gamma$  is total with respect to  $M$  if  $\llbracket A \rrbracket_M \in Dom_M^t(\gamma)$ .*

**Proposition 9.**

- (1) *If a non-logical constant  $A$  of a primitive type is meaningful with respect to a model  $M$  of  $L$ , then  $A$  is total, i.e. if  $A \in Con(\gamma)$  where  $\gamma \in PT$ , and  $A \in Cat_{mf}^M$ , then  $\llbracket A \rrbracket_M \in Dom_M^t(\gamma)$ .*
- (2) *If  $A \in Cat(\langle \alpha, \beta \rangle)$  and  $B \in Cat(\alpha)$  are total with respect to  $M$ , then  $A(B)$  is total with respect to  $M$ .*

**Definition 13.**

- (1) *If  $\approx, \approx'$  are synonymies for  $L$ , we say that  $\approx'$  extends  $\approx$  (or it is an extension of  $\approx$ ) if the field of  $\approx'$  includes that of  $\approx$  and the two synonymies are compatible.*
- (2) *If  $M_1, M_2$  are models of  $L$ , we say that  $M_2$  extends  $M_1$  (or that it is an extension of  $M_1$ ) if  $\llbracket A \rrbracket_{M_2} = \llbracket A \rrbracket_{M_1}$  for all  $A \in Cat_{mf}^{M_1}$ .*
- (3) *If  $M_1, M_2$  are models of  $L$ , we write  $M_2 \geq M_1$  to mean that  $\approx_{M_2} \supseteq \approx_{M_1}$ .*

**Remark 6.** *If  $M_2 \geq M_1$ , then the domain of  $M_2$  includes that of  $M_1$ , but within that domain,  $M_1$  may make more distinctions than  $M_2$  does.*

**Proposition 10.** *The models  $M_1$  and  $M_2$  of  $L$  are equivalent if and only if both  $M_2 \geq M_1$  and  $M_1 \geq M_2$ .*

**Proposition 11.** *If  $M_2$  extends  $M_1$ , then  $M_2 \geq M_1$ . (In this case  $M_2$  makes exactly the same distinctions in the field of  $M_1$  as  $M_1$  does.)*

**Proposition 12.** *A total model is maximal in the sense that all of its extensions are equivalent to it.*

**Proposition 13.** *A total model  $M$  of  $L$  is minimal in the sense that there is no total model  $M'$  such that  $M$  extends  $M'$  and  $M$  and  $M'$  are not equivalent.*

**Corollary 2.** *If a total model  $M$  extends  $M'$  such that  $M$  and  $M'$  are not equivalent, then  $M'$  is a partial model of  $L$ .*

## 5 TARSKIAN MODELS

In Section 4 we investigated the properties of models by means of their generated synonymies. In his well-known paper (Tarski 1936/1983) Tarski introduces a new classification. The classification and therefore the associated synonymy is—at least in some cases—between syntactic synonymy and synonymies generated by non-degenerate models of our language.

**Definition 14.** *If  $L$  is a type-theoretical language,  $M$  is a model of  $L$  and  $A, B$  are well-formed expressions (or grammatical terms, i.e.  $A, B \in \text{Cat}$ ), then we say that  $A, B$  belong to the same semantic category with respect to  $M$  (they have the same  $M$ -category), in symbols  $A \sim_M B$ , if for every expression  $C$  ( $\in \text{Cat}$ ) and a variable  $\tau$  ( $\in \text{Var}$ )*

$$(\lambda\tau C)(A) \in \text{Cat}_{mf}^M \Leftrightarrow (\lambda\tau C)(B) \in \text{Cat}_{mf}^M.$$

In a very general sense, the next proposition has been mentioned by Tarski. In our case it sounds as follows:

**Proposition 14.** *If  $M$  is a (total or partial) model of  $L$ , then  $\sim_M$  is a synonymy with the field of  $\text{Cat}$ .*

**Theorem 4.**  *$A \sim_M B \Rightarrow A \cong_L B$  (and so  $\cong_L \supseteq \sim_M$ ), where  $M$  is a (total or partial) model of  $L$ .*

**Corollary 3.** *If  $A, B$  are well-formed but not meaningful expressions with respect to a partial model  $M$ , i.e.  $A, B \in \text{Cat} \setminus \text{Cat}_{mf}^M$ , then*

$$A \sim_M B \Leftrightarrow A \cong_L B$$

By means of the notion of semantic category, Tarski lays down a very important principle called the first principle of the theory of semantic categories,<sup>13</sup> which is, as he says, very natural “from the standpoint of ordinary usage of language” (Tarski 1936/1983, 216.). In our terminology the principle sounds informally as follows:

<sup>13</sup>Its original version can be found in (Tarski 1936/1983, 216.).

[The first principle of the theory of semantic categories:] Two expressions of our language have the same semantic category if there is an expression of our language such that it produces meaningful expressions when combined with them.<sup>14</sup>

The following definition formulates the first principle of the theory of semantic categories formally, and gives the notion of a Tarskian model:

**Definition 15.** *We say that the model  $M$  of  $L$  is Tarskian if it is the case that if there is a meaningful expression  $C$  and a variable  $\tau$  such that  $(\lambda\tau C)(A)$  and  $(\lambda\tau C)(B)$  are both meaningful, then  $A$  and  $B$  have the same  $M$ -category.*

**Remark 7.** *A model  $M$  of  $L$  is Tarskian if and only if it fulfills Tarski's first principle of the theory of semantic categories.*

**Theorem 5** (Characteristic theorem of Tarskian models). *The model  $M$  of  $L$  is Tarskian, if and only if the synonymies  $\sim_M$  and  $\cong_L$  are equivalent, i.e.  $\sim_M$  equals  $\cong_L$ .*

**Remark 8.** *According to the Characteristic theorem of Tarskian models 5, all Tarskian models of  $L$  have the same system of semantic categories and this system is equivalent to the system of syntactic categories.*

**Proposition 15.** *If the model  $M$  of  $L$  is total, then the synonymies  $\sim_M$  and  $\cong_L$  are equivalent, i.e.  $\sim_M$  equals  $\cong_L$ .*

**Theorem 6.** *If  $M$  is a total model of  $L$ , then  $M$  is Tarskian.*

**Corollary 4.** *Non-Tarskian models are partial.*

## 6 HUSSERLIAN MODELS

In Section 5 we dealt with the connection between syntactic and semantic categories. The next step we have to take is the investigation of the bridge between the system of semantic categories and the classification generated by the equivalence relation  $\approx_M$ .

**Definition 16.**

- (1) *Let  $M_1, M_2$  be models. We say that  $M_1$  and its generated synonymy  $\approx_{M_1}$  are  $M_2$ -Husserlian if  $A \approx_{M_1} B \Rightarrow A \approx_{M_2} B$  for all  $A, B \in \text{Cat}$ .*
- (2) *We say that a model  $M$  of  $L$  is Husserlian if it is  $M$ -Husserlian. (That is  $A \approx_M B \Rightarrow A \sim_M B$  for all  $A, B \in \text{Cat}$ .)*
- (3) *We say that a model  $M (= \langle G, \varrho, v \rangle)$  of  $L$  is strictly Husserlian if  $M' (= \langle G, \varrho, v' \rangle)$  is Husserlian for all assignments  $v'$ .*
- (4) *We say that the generated synonymy  $\approx_M$  of a model  $M$  is Husserlian (strictly Husserlian) if the model  $M$  is Husserlian (strictly Husserlian).*

<sup>14</sup>A version of the principle is quoted by (Hodges 2001b, 11.).

The notion of a Husserlian model creates a connection between generated synonymy and  $M$ -category. It requires that two expressions with the same semantic value with respect to  $M$  have to belong to the same  $M$ -category, and so according to Theorem 4 they have to have the same type. More precisely:

**Proposition 16.** *If a model  $M$  of  $L$  is Husserlian and  $A \approx_M B$  for some  $(A, B \in \text{Cat})$ , then  $A \cong_L B$  i.e. there is a  $\gamma \in \text{TYPE}_{PT}$  such that  $A, B \in \text{Cat}(\gamma)$ .*

**Corollary 5.** *If a model  $M$  of  $L$  is Husserlian, then  $\cong_L \supseteq \approx_M$ , i.e.  $M_L \geq M$ .*

**Corollary 6.** *Let  $M_1, M_2$  be models of  $L$ . If  $M_1$  is  $M_2$ -Husserlian, then it is  $M_L$ -Husserlian.*

**Theorem 7.** *Let  $M$  be a Tarskian model of  $L$ . The model  $M$  is Husserlian if and only if  $\cong_L \supseteq \approx_M$ , i.e.  $M_L \geq M$ .*

**Corollary 7.** *Let  $M$  be a total model of  $L$ . The model  $M$  is Husserlian, if and only if  $\cong_L \supseteq \approx_M$ , i.e.  $M_L \geq M$ .*

Law of replacement 2 says that an expression can substitute for another one without changing the semantic value of the compound expression, if the semantic value of the first expressions equals that of the second one. In the law, there is a special condition usually regarded as not too important. The condition requires that the two expressions have to belong to the same syntactic category. Without supposing it, the law of replacement holds only in Husserlian models. That is why I said that Law of replacement 2 is only a restricted version of the substitutivity principle (see in Section 3), a version of the principle of compositionality. Its unrestricted type-theoretical formulation is the following Husserlian law of replacement.

**Theorem 8** (Husserlian law of replacement). *If  $A, B, C \in \text{Cat}$  and  $M$  is a Husserlian model of  $L$ , then*

$$\llbracket B \rrbracket_M = \llbracket C \rrbracket_M \Rightarrow \llbracket A \rrbracket_M = \llbracket A[C \downarrow B] \rrbracket_M.$$

**Theorem 9** (Conversion of Husserlian law of replacement). *If for all  $A, B, C \in \text{Cat}$*

$$\llbracket B \rrbracket_M = \llbracket C \rrbracket_M \Rightarrow \llbracket A \rrbracket_M = \llbracket A[C \downarrow B] \rrbracket_M,$$

*then  $M$  is a Husserlian model of  $L$ .*

*Proof.* The proof is indirect. Suppose that the model  $M$  is not Husserlian. Then there are  $B, C \in \text{Cat}$  such that  $B \approx_M C$  ( $\llbracket B \rrbracket_M = \llbracket C \rrbracket_M$ ) and  $B \not\approx_M C$ . Therefore there is some  $D \in \text{Cat}$ ,  $\tau \in \text{Var}$ , such that  $(\lambda\tau D)(B) \in \text{Cat}_{mf}^M$  and  $(\lambda\tau D)(C) \notin \text{Cat}_{mf}^M$ . According to Law of replacement 2, it is impossible that  $B, C \in \text{Cat}(\gamma)$  for some  $\gamma \in \text{TYPE}_{PT}$  because in contrary  $\llbracket (\lambda\tau D)(B) \rrbracket_M = \llbracket (\lambda\tau D)(C) \rrbracket_M$ . Therefore there are  $\alpha, \beta \in \text{TYPE}_{PT}$  such that  $\alpha \neq \beta$  and  $B \in \text{Cat}(\alpha)$ ,  $C \in \text{Cat}(\beta)$ . Let  $A = '(\lambda\xi\xi)(B)'$  where  $\xi \in \text{Var}(\alpha)$ .  $A \in \text{Cat}$  and  $A[C \downarrow B] \notin \text{Cat}$  and so  $\llbracket A \rrbracket_M \neq \llbracket A[C \downarrow B] \rrbracket_M$ .  $\square$

**Definition 17.** A model  $M$  of  $L$  fulfills the substitutivity principle if for all  $A, B, C \in \text{Cat}$

$$\llbracket B \rrbracket_M = \llbracket C \rrbracket_M \Rightarrow \llbracket A \rrbracket_M = \llbracket A[C \downarrow B] \rrbracket_M.$$

The next theorem shows that the substitutivity principle is a strong version of the principle of compositionality. In our theoretical framework all models are compositional, but a model fulfills the substitutivity principle if and only if it is Husserlian.

**Theorem 10** (Characteristic theorem of Husserlian models). A model  $M$  of  $L$  is Husserlian if and only if it fulfills the substitutivity principle.

**Definition 18.** A model  $M$  of  $L$  is strongly compositional if it fulfills the substitutivity principle.

**Remark 9.** Characteristic theorem of Husserlian models 10 says the property of being strongly compositional is equivalent to being Husserlian.

**Corollary 8.** If  $M$  is a Tarskian model of  $L$  and  $\cong_L \supseteq \approx_M$ , then it fulfills the substitutivity principle.

**Theorem 11.** A model  $M$  of  $L$  is strictly Husserlian if and only if the sets  $\text{Dom}_M(\gamma)$  ( $\gamma \in PT$ ) are pairwise disjoint ones.

*Proof.* I have to note that the sets  $\text{Dom}_M(\gamma)$  ( $\gamma \in PT$ ) are pairwise disjoint ones if and only if the sets  $\text{Dom}_M(\gamma)$  ( $\gamma \in TYPE_{PT}$ ) are pairwise disjoint ones.

At first we prove that if  $M (= \langle G, \varrho, v \rangle)$  is strictly Husserlian, then the sets  $\text{Dom}_M(\gamma)$  ( $\gamma \in TYPE_{PT}$ ) are pairwise disjoint ones. The proof is indirect. Suppose that  $M$  is strictly Husserlian and there is a semantic value  $u$  such that  $u \in \text{Dom}_M(\alpha) \cap \text{Dom}_M(\beta)$  where  $\alpha \neq \beta$ . Let  $\tau_1 \in \text{Var}(\alpha)$   $\tau_2 \in \text{Var}(\beta)$  and  $v'$  be an assignment such that  $v'(\tau_1) = u = v'(\tau_2)$ . If  $M' = \langle G, \varrho, v' \rangle$ , then  $\llbracket \tau_1 \rrbracket_{M'} = \llbracket \tau_2 \rrbracket_{M'}$  but  $\tau_1 \not\cong_L \tau_2$  and according to Proposition 16  $M'$  is not Husserlian. So  $M$  is not strictly Husserlian.

Secondly it is enough to prove that if  $M$  is a model of  $L$  and the sets  $\text{Dom}_M(\gamma)$  ( $\gamma \in TYPE_{PT}$ ) are pairwise disjoint ones, then  $M$  is Husserlian. The proof is indirect. Suppose that  $A \approx_M B$  and  $A \not\approx_M B$  where  $A, B \in \text{Cat}_{mf}^M$ . Then  $A \not\cong_L B$ , and so there are  $\alpha, \beta \in TYPE_{PT}$  such that  $A \in \text{Cat}(\alpha)$ ,  $B \in \text{Cat}(\beta)$  and  $\alpha \neq \beta$ . According to Proposition 1  $\llbracket A \rrbracket_M \in \text{Dom}_M(\alpha)$ ,  $\llbracket B \rrbracket_M \in \text{Dom}_M(\beta)$ . Since  $\llbracket A \rrbracket_M = \llbracket B \rrbracket_M$ ,  $\text{Dom}_M(\alpha) \cap \text{Dom}_M(\beta) \neq \emptyset$ .  $\square$

**Definition 19.** A (total or partial) frame  $G$  is strictly Husserlian if the sets  $\text{Dom}_G(\gamma)$  ( $\gamma \in PT$ ) are pairwise disjoint ones.

**Corollary 9.** If  $M$  is a model on a strictly Husserlian frame then the model  $M$  of  $L$  is strictly Husserlian.

**Corollary 10.** *The degenerate model  $M_L$ , which generates the syntactic synonymy  $\cong_L$ , is strictly Husserlian, and so the synonymy  $\cong_L$  is strictly Husserlian.*

**Theorem 12.** *A model  $M$  is Husserlian if and only if there is a strictly Husserlian model  $M'$  such that  $M' \geq M$ .*

*Proof.* According to Corollary 10, the model  $M_L$  is strictly Husserlian. If  $M$  is Husserlian, then according to Corollary 5,  $M_L \geq M$ .

Let  $M'$  be a strictly Husserlian model such that  $M' \geq M$ . Then according to Corollary 5  $M_L \geq M'$  and so  $M_L \geq M$ . It means that if  $A \approx_M B$ , then  $A \cong_L B$  i.e. there is  $\gamma \in TYPE_{PT}$  such that  $A, B \in Cat(\gamma)$ . Therefore  $(\lambda\tau C)(A) \in Cat$  if and only if  $(\lambda\tau C)(B) \in Cat$  for any  $C \in Cat$  and  $\tau \in Var$ . According to Law of replacement 2  $\llbracket (\lambda\tau C)(A) \rrbracket_M = \llbracket (\lambda\tau C)(B) \rrbracket_M$  and so  $(\lambda\tau C)(A) \in Cat_{mf}^M \Leftrightarrow (\lambda\tau C)(B) \in Cat_{mf}^M$ , i.e.  $A \sim_M B$ .  $\square$

## REFERENCES

- Church, A., 1940, A formulation of the simple theory of types. *Journal of Symbolic Logic* 5. 56–68.
- Dunn, J. M. and G. M. Hardegree, 2001, *Algebraic Methods in Philosophical Logic*, Vol. 41 of *Oxford Logic Guide*. New York, Oxford University Press.
- Frege, G., 1884/1980, *The Foundations of Arithmetic. A Logic-Mathematical Enquiry into the Concept of Number*. Oxford, Basil Blackwell, second revised edition. Translated by J. L. Austin, from *Grundlagen der Arithmetik. Eine logisch-matematisch Untersuchung über den Begriff der Zahl*. Breslau, W. Koebner.
- Frege, G., 1879/1997, *Begriffsschrift*, a formula language of pure thought modelled on that of arithmetic. In M. Beaney (ed.), *The Frege Reader*. Oxford, Blackwell. 47–78. Selections (Preface and part I). Translated by M. Beaney from *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle, L. Nebert.
- Hodges, W., 2001a, A context principle. Manuscript.
- Hodges, W., 2001b, Formal features of compositionality. *Journal of Logic, Language and Information* 10. 7–28.
- Mihálydeák, T., 2010, On Tarskian models of general type-theoretical languages. In C. Drossos, P. Peppas and C. Tsinakis (eds.), *Proceedings of the 7th Panhellenic Logic Symposium*. Patras, Patras University Press. 127–131.
- Ruzsa, I., 1989, *Logikai szintaxis és szemantika* (in Hungarian). Budapest, Akadémiai Kiadó.
- Ruzsa, I., 1991, *Intensional Logic Revisited*. Budapest, published by the author.
- Ruzsa, I., 1997, *Introduction to Metalogic*. Budapest, Áron Publishers.
- Szabó, Z. G., 2000, Compositionality as supervenience. *Linguistics and Philosophy* 23. 475–505.
- Tarski, A., 1936/1983, The concept of truth in formalized languages. In J. Corcoran (ed.), *Logic, Semantics, Metamathematics*. Indianapolis, Hackett Publishing, second edition. 152–278.
- Thomason, R. H., 1999, Type theoretic foundations of context. Part 1: Contexts as complex type-theoretic objects. In P. Bouquet, L. Serafini, P. Brézillon, M. Benerecetti, and F. Castellani (eds.), *Modeling and Using Contexts: Proceedings of the Second International and Interdisciplinary Conference, CONTEXT'99*. Berlin, Springer-Verlag. 352–374.
- Thomason, R. H., 2001, Contextual intensional logic: type-theoretic and dynamic considerations. Manuscript.

## Ruzsa on Quine's Argument Against Modal Logic

**Abstract.** Through the 1970s and 1980s—the days when ELTE philosophy was named Marxism–Leninism—Imre Ruzsa prepared logic books and articles with sharp, comprehensive, up-to-date surveys of the most recent international developments in logic and the philosophy of language. For decades to come, the chapters of his *Classical, Modal and Intensional Logic* would be just about the only Hungarian-language sources available on W. V. O. Quine's famous argument against modal logic, on Saul Kripke's modal semantics that seemed to bypass the Quinean objections, and on Kripke's arguments about the semantics of natural language: that proper names are rigid designators. Based primarily on John Burgess's subsequent work, we can complete the picture of modal logic that Ruzsa painted in his survey by shedding light on additional important connections: crucial links not so much between Quine's argument and Kripke's formal work (as Ruzsa and others had thought), but instead between the Quinean argument and Kripke's thesis about proper names being rigid designators.

Various stripes of modality—senses of 'must' and 'can', necessity and possibility—are traditionally distinguished by logicians, linguists, and philosophers. Let us list a couple of them:

- Deontic modality—what is necessary/possible given laws or norms; that is, what the laws/norms require/permit. For example, "It is necessary (given public transportation regulations) that I buy a ticket to ride the tram"; more colloquially put: "I must buy a ticket to ride the tram".
- Epistemic modality—what is necessary/possible given what is known. For example, "It is necessary (given what I know) that the Opera building is in the next block"; more colloquially put: "The Opera building must be in the next block".



There is also the category of alethic modality, concerning *truth*—what is necessarily and possibly true. Within this, we can draw further distinctions; let us focus on necessity, leaving possibility aside (as is traditionally done):

- Necessary truth as logical truth (sometimes called ‘strict modality’)—truth given some system of logic, in other words, truth given the meanings of the logical vocabulary of a selected system. “I either buy a ticket or I don’t buy a ticket” is an example of a logical truth, for it is true in virtue of what ‘or’ and ‘not’ mean.
- Necessary truth as analytic truth—truth given the meanings of the words contained in the sentence. “All single people are unmarried” is an example of an analytic truth.
- Necessary truth as physical or natural necessity—truth given the laws of physics/laws of nature. “Trams travel slower than the speed of light” is an example of a truth of physics.
- Indeed, in his state-of-the-art 1984 survey volume *Classical, Modal and Intensional Logic* (written in Hungarian), Imre Ruzsa distinguished each of these stripes of modality (Ruzsa 1984, 119–121., 156–160.). What is conspicuously missing from Ruzsa’s (and his contemporaries’) list is yet another sense of necessity within the alethic category: the notion of counterfactual or metaphysical necessity, brought into the limelight by Saul Kripke’s 1970 lecture series “Naming and Necessity” (subsequently published as Kripke 1980):
- Necessary truth as counterfactual (or metaphysical) necessity—truth across all counterfactual circumstances. “Aristotle is (was) human” is a plausible example of a metaphysically necessary truth. Although it is epistemically as well as logically and analytically possible for Aristotle to be a cat, it is not *counterfactually or metaphysically* possible that he is a cat.

Ruzsa’s *Classical, Modal and Intensional Logic* stood alone in various ways, providing just about the only Hungarian-language coverage of numerous landmarks in philosophy of language and logic for almost two decades:

- (I) W. V. O. Quine’s arguments against modal logic (1943–1962)
- (II) Kripke’s formal results: semantics for modal logic (1959–1963)
- (III) Kripke on the semantics of natural language, specifically, his theory that proper names are so-called rigid designators. (1970)

As for (I), it was not until 2002 that a collection of Quine’s essays was published in Hungarian, including his definitive formulation of his attack on modal logic “Reference and Modality” (Quine 1953, discussed in detail below). Until then, there were just three articles by Quine available in Hungarian: “Two Dogmas of Empiricism” (Quine 1951/1973) as well as two smaller chapters from Quine’s attacks on modal logic (Quine 1963, 1947 both in Copi–Gould 1964/1985). Ruz-

sa's 30-page section entitled "Modality and Quantification: Logic 'Conceived in Sin'" was thus, for quite some time, *the* Hungarian source to consult on Quine's attacks on modal logic (Ruzsa 1984, 164–193).

As for (II), to this day, none of Kripke's formal work has been translated into Hungarian, and Ruzsa's 20-page section entitled "Kripke's Modal Semantics" remains the definitive secondary source to turn to in Hungarian (Ruzsa 1984, 227–248, see also Ruzsa 1988, XX). In addition, Ruzsa went on to develop his own Quine-proof system of modal logic (Ruzsa 1984, 290–345).

As for (III), not until the late 1990s was there any Hungarian coverage or translation of Kripke's *Naming and Necessity* available in Hungarian, apart from Ruzsa's 13-page section on Kripke's rigidity thesis (Ruzsa 1984, 302–315). Kripke argues that proper names like 'Aristotle' exhibit distinctive behavior within a certain rather straightforward kind of modal context: they are rigid designators, that is, they refer to the same individual with respect to every counterfactual situation. The rigidity thesis yields a powerful argument against Frege's descriptivist theory of proper names, which associates proper names with definite descriptions—such as 'the teacher of Alexander the Great'—that are non-rigid: after all, in a counterfactual situation in which someone else taught Alexander the Great, this definite description picks out someone other than Aristotle. The turn of the 20<sup>th</sup>-21<sup>st</sup> centuries brought the Hungarian translation of Kripke's "Identity and Necessity" paper, which also discusses the rigidity thesis (Kripke 1971/2004; see also the brief excerpts collection Kripke 1980/1997). Around the same time, important, albeit brief coverage of the rigidity thesis appeared in (Sainsbury 1997, 85–89) and (Farkas–Kelemen 2002, 135–145). The Hungarian translation of *Naming and Necessity*, along with an 87-page companion article was published fairly recently (Kripke 1980/2007, Zvolenszky 2007). Again, for almost two decades, Ruzsa's 1984 book provided one of very few sources on Kripke's work on the semantics of natural language.

My goal in this short paper is to highlight, beyond (I)–(III), two more aspects of the debate between Quine and Kripke, neither of which have been properly recognized by Ruzsa or his contemporaries:

Supplementing (I): (a) Quine's lasting argument against modal logic, and his challenge to locate an alternative notion of necessity unaffected by his arguments (especially in Quine 1953, 1960, 1963).

Supplementing (III): (b) The role of Kripke's explication of the notion of metaphysical necessity (1970).

These complete the picture painted by Ruzsa's pioneering survey in *Classical, Modal and Intensional Logic*.

From the 1940s through the 1960s, Quine put forth various arguments against modal logic and did not properly distinguish them, which made interpreting him no easy task. One of these arguments—(a)—stands, posing a challenge that was not met until Kripke's observations about counterfactual necessity—(b)—appeared on the scene. Yet this went unrecognized until much later—from the late 1990s, particularly by John Burgess (1997) and Stephen Neale (2000):

(a) Preliminary formulation: Quine's lasting argument: Certain formulas of modal logic lack sense, they cannot be interpreted.

Let us see how we might arrive at such a suspect, uninterpretable formula. Imagine a traveler who knows all too well that the Isonzo river is identical with the Soča. She might still be surprised upon arriving at the river Soča (advertised in brochures as the whitewater rafting paradise of Slovenia), when she learns that it is one and the same river as the Isonzo, the scene of numerous battles in World War I that she had read about in history books. (Indeed, I myself was in for that surprise when travelling to Slovenia: that the Soča is one and the same as the Isonzo constituted a discovery). Thus if we interpret  $\Box$  as, say, epistemic necessity, then (1) is indeed true given what our traveler knows, while (2) is false. Similarly, if we interpret  $\Box$  as analytic necessity—as Quine does—(1) is true given the meanings of the words featured (all of which are familiar to our traveler), while (2) is false (given her subsequent discovery):

- |  |       |
|--|-------|
| (1) It is necessarily true that the Isonzo is identical with the Isonzo.     |       |
| $\Box$ Isonzo = Isonzo   | true  |
| (2) It is necessarily true that the Soča river is identical with the Isonzo. |       |
| $\Box$ Soča = Isonzo   | false |

The truth value assignments for (1) and (2) remain unaltered even if we interpret  $\Box$  as logical necessity, truth in virtue of the meanings of the logical vocabulary. Indeed, it will help our exegesis to introduce the category of *linguistic necessity* to cover both analytic and logical necessity: for both concern truth in virtue of the meanings of certain expressions; the difference is only whether we consider the meanings of all vocabulary items or just the logical ones. Crucially, in formulating his argument (a), Quine's concern was with linguistic necessity (what he called strict necessity), although he rarely made this explicit, especially in his later work.

We can generalize over (1) to arrive at one of the suspect formulas:

- (3) There is a thing  $x$ , such that  $x$  is identical with the Isonzo.  
 $\exists x \Box (x = \text{Isonzo})$

Interpretive trouble ensues: What is this river which, according to (3), is necessarily identical with the Isonzo? According to (1), from which (3) was inferred, it is the Isonzo, that is, the Soča; but to suppose this would conflict with the fact that (2) is false. In a word, to be necessarily [in the linguistic sense] identical

with the Isonzo is not a trait of the river, but depends on the manner of referring to the river. (adapted from Quine 1953, 148)

(3) is an instance of quantifying in, that is, binding the variable  $x$  within the scope of the modal operator  $\Box$  by the quantifier  $\exists x$ , which is outside the scope of  $\Box$ . This is the sort of construction that spells interpretive trouble when it comes to linguistic necessity, according to Quine. He did not think he has given a general argument against quantifying into any modal context whatsoever (as many interpreters at the time thought)—he says this much in the following passage (see also Quine 1963):

What has been said of modality in these pages relates only to strict [that is, linguistic] modality. For other sorts, for example, physical necessity and possibility, *the first problem would be to formulate the notions clearly and exactly*. Afterwards we could investigate whether such modalities, like the strict ones, cannot be quantified into without precipitating an ontological crisis. The question concerns intimately the practical uses of language. ... In discussions of physics, naturally, we need quantifications containing the clause 'x is soluble in water', or the equivalent in words; but ... we should then have to admit within quantifications the expression ... 'necessarily if x is in water then x dissolves'. Yet we do not know whether there is a suitable sense of 'necessarily' into which we can so quantify. (Quine 1953, 158-159.; emphasis added)

Here, Quine poses a challenge: quantifying in spells interpretive trouble for linguistic notions of necessity; when considering how interpretation would go with alternative notions of necessity (physical necessity, for example), first, those notions should be clarified, then the question of interpreting quantifying in can be raised. Accordingly, we can expand (a):

(a) Quine's lasting argument: When considering the (then-)established notion of necessity, that of linguistic necessity, certain modal logic formulas (those involving quantifying in) lack sense, they cannot be interpreted.

Quine's associated challenge: Clarify an alternative notion of necessity, and if the need for interpreting quantifying in arises with respect to that notion, then check that there is no interpretive trouble there.

In what follows we will unpack Quine's lasting argument (following primarily Burgess 1997), and see how Kripke responds to Quine's associated challenge by bringing in the notion of metaphysical necessity. But before that, let us introduce a preliminary distinction between *de dicto* and *de re* statements:

a *de dicto* ("about the sentence") statement:

(4) Necessarily, all single people are unmarried.

"The following is necessary: all singles are unmarried."

a *de re* (“about the thing”) statement:

(5) All single people are necessarily unmarried.

“All singles bear the modal attribute of being necessarily unmarried.”

Consider, for a moment, the counterfactual sense of necessity. According to it, (4) is true, for in all counterfactual circumstances, everyone who is single is unmarried. Meanwhile, (5) is false: for those who are in fact single might, in an alternative scenario, have gotten married instead—they are not single in all counterfactual situations.

Now we can spell out step by step Quine's request for interpreting quantify-in, this time with linguistic necessity at hand:

*Step 1.* First we need to make sense of the open formula ‘ $\Box (x = \text{Isonzo})$ ’.

*Step 2.* This requires making sense of *de re* modal claims.

*Step 3.* The *de dicto* claims at hand are (1) and (2), and their *de re* counterparts are (1r) and (2r):

(1) $\Box \text{Isonzo} = \text{Isonzo}$	true
(1r) It is true of the Isonzo that it is necessarily identical with the Isonzo.	
$\exists x (x = \text{Isonzo} \ \& \ \Box x = \text{Isonzo})$	?
(2) $\Box \text{Soča} = \text{Isonzo}$	false
(2r) $\exists x (x = \text{Soča} \ \& \ \Box x = \text{Isonzo})$	?

But the notion of linguistic necessity—about truth given the meanings of expressions—provides guidance for interpreting *de dicto* modal claims only; there is no direct guidance for making sense of *de re* modal claims. (For what might that *river* be that is *analytically* or *logically* identical with the Isonzo, given that (1) and (2) differ in truth value?)

*Step 4.* We have two strategies for interpreting (1r) and (2r), but both turn out unacceptable.

*Step 5.* The first strategy for interpreting *de re* modal claims is:

*the unselective strategy:* the *de dicto* statement yields its *de re* counterpart—for any proper name whatsoever.

This yields an unacceptable outcome: we have objects with contradictory properties: the river Isonzo a.k.a. Soča is at once analytically identical with the Isonzo (qua Isonzo) and not analytically identical with it (qua Soča). The cost of avoiding this is high: we have to give up on the idea that the truth of *de dicto* modal claims may in part depend on the words and names used. But linguistic necessity is supposed to be about truth in virtue of the meaning of certain expressions, so this option is unacceptable.

*Step 6.* The second strategy for interpreting *de re* modal claims is:

*the selective strategy:* *de dicto* modal claims yield their *de re* counterparts in selected cases only—with respect to standard names.

For example, if 'Isonzo' counts as a standard name while 'Soča' does not, then we cannot get (2r) from (2). But then we would have to make arbitrary decisions about which natural-language proper name to regard as standard: 'Cicero' or 'Tully'? 'Burma' or 'Myanmar'?

*Step 7.* With linguistic necessity, the standard names featured in the selective strategy lead to an arbitrary form of essentialism:

“Evidently, the reversion to Aristotelian essentialism ... is required if quantification into modal contexts is to be insisted on. An object, of itself and by whatever name or none, must be seen as having some of its traits necessarily and others contingently, despite the fact that the latter traits follow just as analytically from some ways of specifying the object as the former traits do from other ways of specifying it.” (Quine 1953, 155)

In other words, with standard names chosen arbitrarily, we end up with arbitrary choices for what is and what is not analytically true of an object. So the distinction between essential and accidental properties of objects—this is what essentialism is committed to—will be arbitrarily drawn.

For linguistic necessity, this seven-step argument does conclusively show that interpreting *de re* modal claims spells trouble whichever interpretive strategy we follow, making the first half of (a) a lasting argument indeed. The second half of (a), Quine's challenge is: we have (yet) to locate a notion of necessity which allows us to make sense of *de re* modal statements without running into unacceptable consequences. It is to this challenge that (Kripke 1980/2007) provides a response:

(b) Kripke's response to Quine's challenge: with the notion of counterfactual (metaphysical) necessity spelled out, interpreting *de re* modal claims is no longer problematic.

The following seem like plausible candidates for true *de re* modal claims: Cicero was necessarily human, but was only contingently born outside Rome; there is a counterfactual situation in which he was born in Rome, but there is no counterfactual situation in which *he* fails to be human. With this counterfactual notion of necessity at hand, our interpretation of *de re* modal claims is directly given; there is no need for either the selective or the unselective strategy of piggybacking on *de dicto* modal claims.

Ruzsa, along with contemporary commentators of Quine, thought that Quine's argument against modal logic (a, that is) targeted all stripes of modality. Hence, they thought that providing a framework for accommodating formulas with quantifying in—Kripke's formal work from the 1950s and 1960s (for example, Kripke 1963)—suffices to show that quantified modal logic is viable after all. (Indeed, commentators were in a difficult position because alongside his lasting argument, Quine also gave other, more general arguments against inter-

preting quantified modal logic, without properly distinguishing them from one another; for problems raised for some of the general arguments, see for example Kaplan 1986 and Fine 1989, 1990.) Ruzsa and others also considered Quine's charge that quantified modal logic comes with a high price tag—embroilment in essentialism, that is, commitment to a distinction between essential and accidental properties of objects (in Step 7)—to arise for quantified modal logics of all stripes. Yet again, there is a crucial detail to realize about Quine's argumentative strategy: his claim that essentialism is untenable is doubly embedded within his argument: first, it is featured within one of the interpretive strategies for making sense of *de re* modal claims (the one based on standard names); and second, we get an arbitrary, and hence objectionable form of essentialism specifically in the case of linguistic necessity, precisely because of the need to rely on standard names. In short, the lasting argument of Quine's does not claim that across the board, there is a problem with interpreting *de re* modal formulas; nor does it claim that across the board, essentialism is objectionable. And the response for his challenge calling for an alternative notion of modality where the interpretive problem is resolved, is in fact met not in Kripke's formal work, but in Kripke's observations about the semantics of natural language, when, in propounding his rigidity thesis, he also clarified the notion of counterfactual necessity (b, that is). (a) and (b) are then the missing links that complete the otherwise admirably detailed and illuminating picture of state-of-the-art modal logic and modal semantics that Imre Ruzsa relayed to Hungarian readers back in 1984.

## REFERENCES

- Burgess, John, 1997, Quinus ab omni naevo vindicatus. In A. A. Kazmi (ed.), *Meaning and Reference: Canadian Journal of Philosophy Supplement* 23, 25–65.
- Copi, Irving M. and Gould, James A (eds.), 1964/1985, *Kortárs tanulmányok a logikaelmélet kérdéseiről (Contemporary Readings in Logical Theory*. In Hungarian, trans. D. Bánki, B. Dajka, I. Faragó-Szabó, K. G. Havas, L. Hársing, A. Máté, K. Solt and L. Urbán). Budapest, Gondolat. Includes translations of Quine 1947 and 1963.)
- Fine, Kit, 1989, The problem of *de re* modality. In J. Almog, J. Perry and H. Wettstein (eds.), *Themes from Kaplan*. Oxford, Oxford University Press, 197–272.
- Fine, Kit, 1990, Quine on Quantifying in. In C. A. Anderson and Joseph Owens (eds.), *Proceedings of the Conference on Propositional Attitudes*. Stanford: CSLI, 1–26.
- Kaplan, David, 1986, Opacity. In L. E. Hahn and P. A. Schilpp (eds.), *The Philosophy of W. V. Quine*. La Salle, IL, Open Court, 229–289.
- Kripke, Saul, 1963, Semantical considerations in modal logic. *Acta Philosophica Fennica* 16, 83–94.
- Kripke, Saul, 1971/2004, Azonosság és szükségszerűség (Identity and Necessity. in Hungarian, trans. F. Csaba). In K. Farkas and F. Huoranszki (eds.), *Modern metafizikai tanulmányok*. Budapest: ELTE Eötvös, 39–68.

- Kripke, Saul, 1980/1997, Névadás és szükségszerűség (részletek az első és a második előadásból). (Short excerpts from *Naming and Necessity* in Hungarian, trans. M. Wappel). *Helikon* 43, 410–426.
- Kripke, Saul, 1980/2007, *Megnevezés és szükségszerűség*. (*Naming and Necessity*. In Hungarian, trans. T. Bárány). Zs. Zvolenszky (ed.), Budapest. Akadémiai.
- Farkas, Katalin and Kelemen, János, 2002, *Nyelvefilozófia*. (*Philosophy of Language*. In Hungarian.) J. Bárdos (ed.), Budapest, Áron.
- Neale, Stephen R., 2000, On a milestone of empiricism. In P. Kotatko and A. Orenstein (eds.), *Knowledge, Language and Logic: Questions for Quine*. Dordrecht, Kluwer 237–346.
- Quine, Willard V., 1947, The problem of interpreting modal logic. *Journal of Symbolic Logic* 12, 43–48. (Hungarian translation in Copi–Gould 1964/1985.)
- Quine, Willard V., 1951/1973, Az empirizmus két dogmája. (Two Dogmas of Empiricism. In Hungarian, trans. I. Faragó-Szabó.) *Magyar Filozófiai Szemle* 47, 225–239.
- Quine, Willard V., 1953/2002, Reference and Modality. (Reference and modality. In Hungarian, trans. E. Boldizsár.) In Quine 2002, 225–251.
- Quine, Willard V., 1960, *Word and Object*. Cambridge, MA, MIT Press.
- Quine, Willard V., 1963, (1962). Reply to Professor Marcus. In M. Wartofsky (ed.), *Boston Studies in the Philosophy of Science*. Dordrecht, D. Reidel, 97–104. (Hungarian translation in Copi–Gould 1964/1985.)
- Quine, Willard V., 2002, *A tapasztalattól a tudományig*. (*From Experience to Science*. Gábor Forrai (ed.) Budapest, Osiris.
- Ruzsa, Imre, 1984, *Klasszikus, modális és intenzionális logika*. (*Classical, Modal and Intensional Logic*. In Hungarian.) Budapest, Akadémiai.
- Ruzsa, Imre, 1988, *Logikai szintaxis és szemantika* vols. 1–2. (*Logical Syntax and Semantics*. In Hungarian.) Budapest, Akadémiai.
- Sainsbury, Mark, 1995/1997, Filozófiai logika. (Philosophical logic. In Hungarian, trans. K. Farkas.) In A. C. Grayling (ed.), *Filozófiai kalauz*. (*Philosophy: A Guide through the Subject*. In Hungarian.) Budapest, Akadémiai, 73–140.
- Zvolenszky, Zsófia, 2007, *Megnevezés és szükségszerűség – Négy évtized távlatában* (Four decades of *Naming and Necessity*. In Hungarian.) In Kripke 1980/2007, 151–218.



# PHILOSOPHICAL LOGIC AND ITS HISTORY

---

ANNA BROŻEK

## On the so-called embedded questions

**Abstract.** The analysis of dependent questions plays an important role in the general theory of questions. Dependent questions are expressions which are parts of compound questions and are isomorphic with some independent questions (scil. questions *sensu stricto*). One may meet the tendency to explicate the sense of independent questions by the sense of dependent ones, e.g. the sense of questions such as “Where is Budapest situated?” is explicated by the sense of sentences such as “A knows where Budapest is situated”, where the second contains the first as a part. The analysis of dependent questions is often the point of departure for constructing set-theoretical or possible worlds semantics for independent questions. In my opinion, these tendencies are abortive and lead to irrelevant explications of the sense of questions *sensu stricto*. But on the other hand, semiotic functions of the so-called dependent questions as parts of compound expressions require deeper analysis. My paper contains a proposal of such an analysis.

### 1 SUPPOSITIONS

The following are the points of departure for my paper:

(A) Following Kazimierz Ajdukiewicz (1956/1978), I distinguish stating from expressing; for instance, if a person  $P$  utters the sentence ‘ $p$ ’ then this sentence *states* the occurrence of a certain state of affairs (namely that  $p$ ), and *expresses* the conviction of  $P$  that  $p$ .

(B) I adopt the concept of name introduced by Stanisław Leśniewski: “a name” means an expression that can occur as a subject of a subject-predicate sentence, or as part of a predicate in a subject-predicate sentence of the form “... is ...”; this means that both “Imre Ruzsa” and “the inventor of the system of intensional logic with semantic value gaps” are names, since the first is the subject and the second is part of the predicate within the sentence “Imre Ruzsa is the inventor of the system of intensional logic with semantic value gaps.”

(C) I distinguish designating from denoting;

- (a) a name  $N$  designates an object  $A$  iff  $N$  can be truly predicated about  $A$  or  $A$  can be indicated by  $N$ ;
- (b) the denotation of  $N$  is the set of all things designated by  $N$ —its designata.

## 2 *SENSU STRICTO* QUESTIONS VS. NOMINALIZED QUESTIONS

The communicative sense of every *sensu stricto* question is composed of three elements: cognitive, incognitive and volitional. For instance, if a person  $P$  asks seriously:

(1) Where was Imre Ruzsa born?

then  $P$  expresses:

- (a) that  $P$  is convinced that Imre Ruzsa was born somewhere;
- (b) that  $P$  does not know where Imre Ruzsa was born;
- (c) that  $P$  wants to know where Imre Ruzsa was born.

These three components of sense distinguish questions as a specific class of expressions.

It is convenient to describe this situation using the metaphor of a picture of a situation. A person seriously uttering a question has a mental picture of a situation with an epistemic gap, and wants to fill that gap.

Questions have the following general form:

(2)  $?x (Fx)$

i.e. “For which  $x$  it is a fact that  $Fx$ ?” This scheme was first proposed by Kazimierz Ajdukiewicz in 1923 (13 years before Rudolf Carnap, who is usually credited with introducing it). Later, Tadeusz Kubiński (1970) used Ajdukiewicz’s formulation in the construction of his systems of erotetic logic (*i.e.* his logic of questions), noting the analogy between the role of the questionmark in (2) and the role of quantifiers in declarative sentences.

Example (1) is a complete question; in what follows, I shall use examples of such questions only. However, my remarks may easily be expanded to other kinds of questions (selective and confirmative ones), since all questions—after appropriate preparation—come under the scheme (2). The only difference is in the scope of the unknown and the way of defining it.

## 3 REDUCTIONS

Logical theories of questions usually simplify the sense of questions: they reduce it to an exclusively cognitive, exclusively incognitive, or exclusively procognitive element. In my opinion, none of these three elements should be omitted when

constructing a materially adequate theory of questions.

One could argue that since I am able to list the elements of the sense of questions, I should agree that questions of the form (1) uttered by me may be reduced to the conjunction of declarative sentences of the form:

(3)  $(I \text{ know that } \exists x (Px)) \wedge \sim \exists x (I \text{ know that } Px) \wedge \forall x (Px \rightarrow I \text{ want to know that } Px)$

where the variable  $x$  ranges over (names of) places and  $P$  is the property of being-a-place-of-birth-of-Imre-Ruzsa. There are at least two reasons why (3) is not an adequate paraphrase of (1). Firstly, the expression “I want to know that  $Px$ ”, being a component of (3), is semantically defective. We encounter it sometimes in ordinary situations, but only in the sense “I want you to tell me that  $Px$ ”, which is not of course the proper sense of “to know”. Secondly, the sense of (3) is essentially different from the sense of (1): one may express this difference by saying that (3) states what (1) expresses. To state that one possesses experiences motivating one to pose a question is not the same as actually to pose that question. One may experience everything that is stated in (3) without asking (1) at all.

Both these reasons for rejecting the paraphrase (3) are important in the case of so-called embedded questions.

#### 4 EMBEDDED QUESTIONS: MISUNDERSTANDINGS

Let us use the term “embedded questions” to denote the set expressions isomorphic to *sensu stricto* questions but being proper parts of declarative sentences (not merely quoted in them).

The first misunderstanding connected with embedded questions is that one may reduce *sensu stricto* questions to declarative sentences containing embedded questions (*scil.* that one may explicate the sense of *sensu stricto* questions through the sense of embedded ones).

Such reductions are proposed, e.g., in the imperative-epistemic tradition in the theory of questions where exclusively embedded questions are used—as a certain step—within paraphrases of questions (see Åquist 1965).

For instance, at the point of departure in one of the versions of this concept, (1) is paraphrased as follows:

(4) Let it be the case that I know where Imre Ruzsa was born.

In the next step, sentences like (4) are paraphrased in such a way that they do not contain embedded questions: they are equal to sentences containing the predicate “know that” which has been well analyzed by logicians, e.g.:

(5)  $\forall x (\text{Imre Ruzsa was born in } x \rightarrow \text{let it be the case that I know that Imre Ruzsa was born in } x)$ .

The paraphrase (5) ignores the aforementioned distinction between expressing and stating or describing: (5) describes components of the sense expressed in (1). In addition, paraphrase (5) violates our linguistic intuitions by introducing the expression “I want to know that” which (as was previously observed) seems to be incorrect.

The second misunderstanding connected with the concept of embedded questions is that the point of departure of so-called erotetic semantics should be (or at least could be) the semantics of embedded questions (see Lahiri 2002). Such a view is incorrect simply because embedded questions are not *sensu stricto* questions. In fact, according to such an approach, one constructs nothing over and above a semantics of declarative sentences containing embedded questions.

However, the problem of the sense of embedded questions is intriguing.

## 5 EMBEDDED QUESTIONS AS NAMES

Let us now analyze the problem of what the function of embedded questions in declarative sentences is. Consider the sentence:

(6) Ferenc knows where Imre Ruzsa was born.

This contains as a component an expression isomorphic to (1), i.e. the expression “Where was Imre Ruzsa born?” (the only and usually ignored difference is inversion). Thus we notice an analogy between (6) and the sentence:

(7) Ferenc knows that Imre Ruzsa was born in Budapest.

since (7) contains the sentence:

(8) Imre Ruzsa was born in Budapest.

as a component. There are many possible analyses of (7); the most popular of these takes the expression “knows that” as the main predicate with two arguments: name-argument and sentence-argument. In another interpretation—a less popular but more accurate one—the predicate “know” takes two name-arguments, the second argument being a name of a suitable situation. In Polish, one may even say:

(9) Ferenc wie to, że Imre Ruzsa urodził się w Budapeszcie.

and see explicitly the «*reificator*» “to, że” (Eng. “that”) of the sentence occurring after “wie” (Eng. “knows”). In word-for-word translation, the start of sentence (9) has the form “Ferenc knows this [fact] that...”.

Let us analyze sentence (6) analogously. We accept that the predicate “know” in (6) possesses two name-arguments (and not name-argument and question-argument). Again, we may say in Polish:

(10) Ferenc wie to, gdzie urodził się Imre Ruzsa.

The initial phrase of (10)—“Ferenc wie to, gdzie”—has the structure of the type “Ferenc knows this [fact] where...”. It is hypothesized that embedded questions are always preceded by an explicit or implicit reificator (or nominalizator). There

are several arguments in favor of such an analysis of (at least some) embedded questions.

Firstly, the following expression, being a paraphrase of (6), possesses explicitly two name-arguments:

(11) Ferenc knows the place of Imre Ruzsa's birth.

Secondly, in Polish (and probably some other languages), embedded questions which occur at the beginning of the sentence possess an obligatory reificator, e.g.

(12) To, gdzie urodził się Imre Ruzsa, ciągle pozostawało dla Ferenc'a tajemnicą.

Maybe the lack of reificator *inside* the sentence is caused only by the specific connectivity of some verbs. In English, an explicit reificator of the type "this [fact]" does not appear:

(13) Where Imre Ruzsa was born was still a mystery for Ferenc.

But the position of the embedded question at the beginning of the sentence, and its specific word order, make its name-like character more clear.

Thirdly, embedded questions do not perform the communicative function of questions (mentioned in section 1). The person uttering (6) does not reveal the desire to fill a gap in a picture of a situation. The situation is similar with embedded sentences. The communicative function of sentences consists in expressing convictions. But a person uttering (7) does not express the conviction that Imre Ruzsa was born in Budapest—only the conviction that Ferenc knows that Imre Ruzsa was born in Budapest.

In what follows, I assume that at least some embedded questions are nominalized questions. I also assume that a question that has undergone nominalization does not perform the same functions as a *sensu stricto* question (just as a nominalized sentence does not perform the same functions as the *sensu stricto* sentence). Nominalized questions are names and—like every name—they have referential functions, *scil.* they designate something. The problem is to say what the designata of nominalized questions are.

## 6 THE DESIGNATA OF NOMINALIZED QUESTIONS

Consider the following sentences containing nominalized sentences:

(14) Ferenc knows that Imre Ruzsa was born in Budapest.

(15) Ferenc was convinced that Imre Ruzsa was born in Budapest.

(16) That Imre Ruzsa was born in Budapest influenced his life.

(17) That Imre Ruzsa was born in Budapest encourages Ferenc to take part in the conference *Logic, Language, Mathematics* devoted to the author of *Modal Logic with Descriptions*.

What does the name  $N$ : “that Imre Ruzsa was born in Budapest” in sentences (14)–(17) refer to? Generally speaking, one usually assumes that nominalized sentences designate states of affairs (i.e. elements of reality or objects abstracted from reality), or judgments (i.e. elements of thoughts or objects abstracted from thoughts). The following are possible approaches to this problem (two uniform and one mixed approach):

- (a) every nominalized sentence refers to situations;
- (b) every nominalized sentence refers to judgments;
- (c) nominalized sentences are ambiguous: in one sense they refer to situations; in the second sense to judgments.

At first glance, it seems that in (14) and (15), the name  $N$  refers to a state of affairs, whereas in (15) and (17) it refers to a judgment. This implies the mixed solution: it is hard to defend any homogeneous one. I omit this problem since it does not relate to the main theme of my investigations.

Let us stress once again: nominalized sentences perform different semiotic functions than non-nominalized sentences do. *Sensu stricto* sentences describe the occurrence of states of affairs and are used to express convictions. Nominalized sentences designate states of affairs or judgments; they are not used to express convictions.

The situation appears similar to the case of nominalized questions: they perform a different semiotic function than *sensu stricto* questions.

Consider the question:

(18) Who was born in Budapest?

Let us keep in mind that somebody who seriously poses such a question possesses a gapped picture of a situation; this situation involves the relation  $\_$  was-born-in $\_$ , with the gap in the first argument, the second being known (it is Budapest). Somebody who seriously utters (18) wants to fill this gap.

Let us call a person’s particular experiences, composed of these three components (cognitive, incognitive and volitional), “inquiries”. An inquiry understood in such a way—as expressed in questions—is a counterpart of the convictions expressed in sentences.

In reality, there are no *gapped* states of affairs. But our pictures of real situations possess gaps. However, questions are correlated with some specific *full* situations—situations which one asks for, pictures of which we aim to possess when we pose questions. Sentences stating the existence of these states of affairs constitute accurate (i.e. true and direct) answers to questions. Let us call states of affairs which are correlates of true answers of a given question “supplementations” of that question. It should be stressed that some questions—in particular, improperly posed questions—do not have supplementations, since they do not possess accurate answers.

A supplementation of a question of the type ‘? $x$  ( $Px$ )’ is identical with such a state of affairs whose occurrence is stated by a true substitution of the formula

' $Px$ '. For instance, the fact that Imre Ruzsa was born in Budapest is one of the supplementations of the question (18) (this question has of course many other supplementations).

Nominalized sentences are not suitable to express conviction; nominalized questions are likewise not suitable to express inquiries. However they are suitable to indicate inquiries or supplementations.

Three possibilities may be considered:

- (a) every nominalized question refers to an inquiry;
- (b) every nominalized question refers to supplementation;
- (c) nominalized questions are ambiguous: in one sense, they refer to inquiries, in another sense, to supplementations.

## 7 INQUIRIES AS CORRELATES OF NOMINALIZED QUESTIONS

Consider the set of famous logicians, the nominalized question "[that] who was born in Budapest", and its role in sentences:

- (19) Ferenc asked [about] who was born in Budapest.
- (20) Ferenc knows who was born in Budapest.
- (21) Ferenc wanted to know who was born in Budapest.
- (22) Who was born in Budapest was a mystery.
- (23) Who was born in Budapest influenced the fate of the city.
- (24) Who Budapest's citizens are proud of depends on who was born in Budapest.

What would it mean to say that the nominalized questions in (19)–(24) refer to inquiries? *Sensu stricto* (not nominalized) questions communicate the desire to fill a gap in the picture of a situation. It seems that in the question-state indicated (by a nominalized sentence), the volitional element is not included; the only indicated elements are the cognitive and incognitive ones. In other words, nominalized sentences designate *gapped* pictures of situations.

Such a solution is implied first of all in contexts in which nominalized questions are arguments of predicates such as "ask", "wonder", "inquire", "guess", etc.

## 8 "FULL" STATES OF AFFAIRS AS CORRELATES OF NOMINALIZED QUESTIONS

In some contexts, it seems that nominalized questions refer to supplementations. In particular, the name "that- $q$ " designates the state of affairs designated by "that  $p$ ", where ' $p$ ' is a true answer to ' $q$ '.

It was observed long ago that in the case of the verb “know”, the following dependencies hold:

(25) If Ferenc knows where Imre Ruzsa was born and Imre Ruzsa was born in Budapest, then Ferenc knows that Imre Ruzsa was born in Budapest.

The same applies in the case of such verbs as “say” (as a synonym of “inform”, and not “utter”), “be surprised”, “it is clear that”. But notice that such a solution (i.e. considering nominalized questions as names of supplementations) can also be applied in cases (19)–(24). For instance, it is a certain state of affairs which is unknown (say, it is a mystery) (example (22)); they are certain facts such that a relation of dependency holds between them (example (24)), etc.

## 9 PRAGMATIC PROPERTIES OF NOMINALIZED SENTENCES AS NAMES OF STATES OF AFFAIRS

Let us agree that the following sentence is true:

(26) Imre Ruzsa was born in Budapest.

Now, consider the name:

(27) [that] Imre Ruzsa was born in Budapest.

Let us agree also that the name (27) designates a certain fact; let us call this fact *f*. Now, consider the following names:

(28) [that] somebody was born in Budapest

(29) [that] Imre Ruzsa was born somewhere

Both these names designate *f*; but (28) designates additionally other states of affairs.

Consider, finally, the following names:

(30) [that] who was born in Budapest

(31) [that] where Imre Ruzsa was born

What are their designata?

If we agree that nominalized questions designate supplementations (in my sense), then the designata of (30) and (31) are the same as in the case of (28) and (29). It seems that (30) designates *f* and other states of affairs, whereas (31) designates only *f*.

It is not surprising that two names designate the same object. But what is the difference between the two types of names?

Note that we use nominalized questions in specific situations, i.e. when we cannot indicate the supplementation precisely or when we do not want to indicate the filling of a gap. This may be easily seen from the following examples:

(32) Ferenc knows who was born in Budapest (but I do not know).

(33) I know who was born in Budapest (but I shall not say).

Moreover, nominalized questions are used when we want to express general dependencies:



(34) How successful our conference is depends on what the weather is like.

(35) Whether I understand Imre Ruzsa's works depends on what the language in which they were published is.

These sentences have a more general sense than sentences with the functor "if-then":

(36) If the weather is dreadful, then our conference will turn out well.

(37) If the language in which Imre Ruzsa's works were published is Hungarian, then I do not understand them.

Again—in contrast with (34) and (35)—in the case of (36) and (37), we do not know or do not reveal consciously what exactly this relation consists in.

## 10 SUMMARY

Let me summarize my views.

- Firstly, embedded questions are not *sensu stricto* questions.
- Secondly, questions in embedded contexts are (explicitly or implicitly) nominalized questions.
- Thirdly, nominalized questions are ambiguous: in one sense, they designate question-states, while in another, they designate supplementations.
- Fourthly, nominalized questions have denotations similar to some nominalized sentences, but they perform more sophisticated pragmatic functions.

## REFERENCES

- Ajdukiewicz, K., 1923, O intencji pytania 'Co to jest *P*?' (On the intention of the question 'What is *P*?). *Ruch Filozoficzny* 8, 152b-153a.
- Ajdukiewicz, K., 1956/1978, Conditional statement and material implication (Okres warunkowy a implikacja materialna.) In: K. Ajdukiewicz, *The Scientific World-Perspective and Other Essays*, Dordrecht, Reidel, 222–238.
- Åquist, L., 1965, *A New Approach to the Logical Theory of Interrogatives*. Upsala, Almqvist & Wiksell.
- Kubiński, T., 1970a, *Wstęp do logicznej teorii pytań. (Introduction to the Logical Theory of Questions.)* Warszawa, PWN.
- Lahiri, U., 2002, *Questions and Answers in Embedded Contexts*. Oxford, Oxford University Press.

GYULA KLIMA

## Natural Logic, Medieval Logic and Formal Semantics

**Abstract.** This paper provides a comparative analysis of the issue of natural logic: the “formalizational approach”, prevalent in contemporary logic, and the “regimentational approach”, prevalent in medieval logic, as exemplified by the 14<sup>th</sup>-century nominalist philosopher, John Buridan. The differences between the two are not as great as they may first appear: a little tweaking of standard quantification theory can take us surprisingly close to Buridan’s logic. However, as the conclusion of the paper points out, there still are some fundamental differences between the resulting “reconstructed Buridanian logic” and Buridan’s logic itself, discussed in detail in the author’s recent monograph.

### NATURAL LANGUAGE AND THE IDEA OF A “FORMAL SYNTAX” IN BURIDAN

The idea of the universality of logic is based on the conviction that despite the immense diversity of human languages, there are certain invariant features of human reasoning, carried out in any natural language whatsoever, that allow the formulation of universal logical laws, applicable to any language. It is precisely for expressing these universal, invariant aspects of human reasoning that in modern logic we construct an artificial language, which is then conceived to serve as a more direct linguistic expression of those invariant conceptual structures that are variously expressed by various natural languages.

But this is not the only possible way to achieve the desired transparency of conceptual structure through the transparency of syntax. The way the 14<sup>th</sup>-century nominalist philosopher, John Buridan (and medieval logicians in general) achieved this was by using, *not* a full-fledged artificial language, but an artificially “regimented” Latin. We can get a nice, yet relatively simple illustration of what this “regimentation” of Latin consists in if we take a closer look at how Buridan introduces the idea that every simple categorical proposition of Latin can be

reduced to the “canonical” subject-copula-predicate form. After briefly stating the division of propositions into categorical and hypothetical, and the description of categorical propositions as those that consist of subject and predicate as their principal parts, Buridan remarks:

... a verb has to be analyzed into the verb ‘is’ as third adjacent, provided that the proposition is assertoric [*de inesse*] and in the present tense [*de praesenti*], and into the participle of that verb, as for example, ‘A man runs’ is to be analyzed into ‘A man is running’, and similarly, ‘A man is’ into ‘A man is a being’.<sup>1</sup>

English speakers might at once notice that the proposed transformation does not always yield equivalent sentences, given the tendency in English to use the simple present tense to signify habitual action as opposed to the continuous present tense, consisting of the copula and the appropriate participle, which is used to express present action. For instance, if I say ‘I smoke’, I may simply want to express that I am a smoker, a person who has the habit of smoking, but this does not mean that I am actually smoking, which would properly be expressed by the sentence ‘I am smoking’. In fact, in accordance with Buridan’s theory of predication, according to which the affirmative copula expresses the identity of the *supposita*, that is, the referents of the terms flanking it, a more appropriate rendering of his proposed transformation would be ‘I am [identical with] someone smoking’.

But Buridan might answer that this is merely a difference in the different syntactical “clues” a different language uses to indicate a different sort of underlying conceptual construction. The simple present tense of English, when it is used to signify habitual action, should then not be analyzed into a participle and a simple assertoric copula, but *perhaps* (somewhat unidiomatically) into a participle and an adverbially modified copula, as in ‘I am usually smoking’,<sup>2</sup> where we just express in the surface syntax of this sentence an adverbial modifier that is unmarked in the simple tense (as is the implicit copula), but which is nevertheless present in the corresponding mental proposition. In any case, it is in this spirit that Buridan answers four questions he raises in connection with the issue of the “canonical form” of categorical propositions:

<sup>1</sup> SD 1.3.2.

<sup>2</sup> Alternatively, one might say that the best explication of ‘I smoke’ expressing the habit is ‘I am a smoker’, where the nominal definition of ‘smoker’ may explicate the habit, as in ‘x is a smoker iff x has the habit of smoking’. But as Buridan often remarks, “examples are not to be verified”, i.e., it does not matter whether we provide “the right analysis” here, as long as it serves to illustrate the point.

But then some questions arise. The first concerns what such a copula signifies. The second is whether that copula is a principal part of a categorical proposition. The third question is whether the proposition 'The one lecturing and disputing is a master or a bachelor' is categorical or hypothetical; for it seems that it is hypothetical, since it has two subjects and two predicates. The fourth question is the same concerning the proposition 'A man who is white is colored'; for it seems that it is hypothetical, since here we have two subjects, two predicates and two copulas; and also because it seems to be equivalent to 'A man is colored, who is white' which is apparently hypothetical.<sup>3</sup>

In his reply, Buridan provides the rationale for the canonical subject-copula-predicate structure in terms of what modern linguists would certainly recognize as "deep structure", and what for Buridan is the conceptual structure of the corresponding mental proposition:

To the first question we should reply that a spoken proposition has to signify a mental proposition [...]. A mental proposition, however, involves a combination of concepts [*complexio conceptuum*], and so it presupposes in the mind some simple concepts, to which it adds a complexive concept, by means of which the intellect affirms or denies one of those [presupposed simple] concepts of the other. Thus, those presupposed concepts are the subject and the predicate in a mental proposition, and they are called the matter of the mental proposition, for they are presupposed by the common form of a proposition, just as matter is presupposed by the substantial form in [the process of] generation. And then it is clear that the subject and the predicate of the spoken proposition signify in the mind the subject and the predicate of the mental proposition. The copula 'is' signifies an affirmative complexive concept, whereas the copula 'is not' signifies a negative complexive concept; and the intellect is unable to form that complexive concept except when it has formed those which are the subject and the predicate, for it is impossible to have the combination [*complexio*] of the predicate with the subject without the predicate and the subject. And this is what Aristotle meant<sup>4</sup> when he said that 'is' signifies a certain composition which cannot be understood without the components.<sup>5</sup>

What fundamentally justifies sticking to the idea of this "canonical form" according to Buridan is that no matter how a mental proposition is expressed

<sup>3</sup> SD 1.3.2.

<sup>4</sup> Aristotle, *On Interpretation*, 1, 16b24.

<sup>5</sup> SD 1.3.2.

in the (“surface”) syntax of a spoken language, the concept of the copula is there in the mental proposition. Therefore, indicating it in the syntax of the spoken proposition merely explicates the presence of the complexive concept of the copula in the corresponding mental proposition. Indeed, this explication is *always* justified because, as Buridan now explains in his answer to the second question, that complexive concept *has to* be present in *any* mental proposition:

To the second question we should reply that the copula is truly a principal part of the proposition, because there could not be a categorical proposition without it; and also because it can be compared to a form of the subject and the predicate, and the form is a principal part of a composite.<sup>6</sup>

Thus, given that the copula is the “formal”, principal part of a categorical proposition, i.e., it is that complexive concept (proposition-forming functor) in the mind without which the concepts corresponding to the terms would not constitute a proposition, it is obvious that no matter how complex those terms and the corresponding concepts are, if they are joined by one copula, then they form one proposition. This is precisely the basis of Buridan’s answer to the third question:

To the third question we should reply that that proposition is categorical; for it does not contain two categoricals, as there is only one copula here; neither are there several subjects, nor several predicates here, for the whole phrase ‘the one lecturing and disputing’ is a single subject [...], namely, a conjunctive subject, and the whole phrase ‘master or bachelor’ is likewise a single predicate, although disjunctive.<sup>7</sup>

As this remark clearly illustrates, Buridan would allow complex terms in either the subject or the predicate positions of otherwise simple, categorical propositions. In fact, given the potentially unlimited complexity of the terms of categorical propositions, these propositions may exhibit a *very* complex structure *within* their terms, despite the simplicity of the general subject-copula-predicate scheme. For it is not just the (iterable) “Boolean” operations of disjunction, conjunction and negation that can yield potentially infinite complexity in these terms, but also the fact that any proposition can be turned into a term (by forming a “that-clause”) or into a determination of a determinable term (in the form of a relative clause). For example, if we take the proposition ‘A man is running’, it can easily be transformed into the term ‘That a man is running’, which can then

<sup>6</sup> SD 1.3.2.

<sup>7</sup> Ibid. Note that in Buridan’s usage, ‘hypothetical’ in this context simply means ‘complex’, as opposed to the widespread modern usage that makes it equivalent to ‘conditional’.

be the subject of another proposition, e.g., 'That a man is running is possible' or a part of another more complex term in another proposition, as in 'That a man is running is believed by Socrates'. Again, taking the proposition 'A man is white', and inserting a relative pronoun after its subject, we get another complex term 'A man who is white', which can then be the subject in the proposition 'A man who is white is colored'.

Now if we look at this proposition in this way, namely, as having a complex subject term built up from a head noun as the *determinable* determined by a relative clause, then it should be obvious why Buridan gives the following answer to the problem raised in connection with this proposition:

To the fourth question we should reply that there is one predicate here, namely, 'colored', which by the mediation of the copula is predicated of the whole of the rest as of its subject, namely, of the whole phrase: 'man who is white'; for the whole phrase: 'who is white' functions as a determination of the subject 'man'. And the case is not similar to 'A man is colored, who is white', for there are two separate predicates here, which are predicated separately of their two subjects, and there is not a predicate here which would be predicated by the mediation of one copula of the whole of the rest. And although these [propositions] are equivalent, they are not equivalent if we add a universal sign. For positing the case that every white man runs and there are many others who do not run, the proposition 'Every man who is white runs' is true, and is equivalent to: 'Every white man runs'; but the proposition 'Every man, who is white, runs' is false, for it is equivalent to: 'Every man runs and he is white'.<sup>8</sup>

Buridan's response to the objection in terms of distinguishing two interpretations of the relative clause indicated by different word order is particularly revealing of his practice of using a "regimented Latin" to make logical distinctions. Indeed, the difference between the syntactical devices used in English and Latin to make the same distinction is also very instructive concerning the advantages and disadvantages of developing logical theory in a "regimented" natural language, as opposed to doing the same using an artificial language, as we usually do nowadays.

Let us take a closer look at the syntax and the semantics of the propositions distinguished here, both in English and in Latin:

- (1) Homo qui est albus est coloratus
- (2) A man who is white is colored
- (3) Homo est coloratus qui est albus
- (4) A man, who is white, is colored

<sup>8</sup> SD 1.3.2.

- (5) *Omnis homo qui est albus currit* ↔ (5') *Omnis homo albus currit*  
 (6) *Every man who is white runs* ↔ (6') *Every white man runs*  
 (7) *Omnis homo currit qui est albus* ↔ (7') *Omnis homo currit et ille est albus*  
 (8) *Every man, who is white, runs* ↔ (8') *Every man runs, and he is white*

Every other line here is the English translation of the Latin of the preceding line. Yet, the syntactical devices by which the Latin and the English sentences bring out the intended conceptual distinction are obviously different (word order vs. punctuation). Nevertheless, the important thing from our present point of view is that these different devices can (and do) bring out the *same* conceptual distinction.

As should be clear, the fundamental difference in all the contrasted cases is whether the relative clause is construed as a *restrictive relative clause*, forming part of the complex subject term, or it is construed as a *non-restrictive relative clause*, making a separate claim referring back to the simple subject of the main clause.

The “regimentation” of the syntax of a natural language, therefore, is the *explication*, and occasionally even the *stipulation*, of *which* syntactical structures of the given language are supposed to convey *which* conceptual constructions. The governing principle of Buridan’s regimentation of his technical Latin seems to be what may be called the *principle of scope-based ordering*. This principle is most clearly at work in the “Polish notation” in modern formal logic (where the order of application of logical connectives is indicated by their left-to-right ordering), but something similar is quite clearly noticeable in Buridan’s rules of logical syntax in general.

To be sure, Buridan never goes as far as to organize Latin according to the rules of a formal syntax in the way a modern artificial language is constructed.<sup>9</sup> And for all his views about the conventionality of language, even he would shy away from re-writing the rules of Latin grammar to fit the requirements of the syntax of an artificial language. Rather, he uses the existing grammatical, structural features of Latin (sometimes stretching, and sometimes bending them a little) to make conceptual distinctions. However, once such a distinction is somehow made, using some such existing syntactical device, Buridan’s regimentation of Latin consists in his insistence on the point that this syntactical device should be consistently regarded as expressing this conceptual distinction, at least when we use language for the purposes of logic (as opposed to, for example, using it to do poetry).

<sup>9</sup> I tried to do this once for a *tiny* fragment of Latin with an explicitly listed finite vocabulary for the purposes of illustration, and even that resulted in an extremely complex, unwieldy system. See (Klima 1991).

## REGIMENTATION VS. FORMALIZATION

Thus, even if doing logic by means of a full-fledged artificial, formal language did not even emerge as a theoretical alternative for Buridan, given the fact that in our time this is the dominant approach to logic, we should pause here a little to reflect on the theoretical as well as the practical advantages and disadvantages of these two different approaches.

One apparent disadvantage of Buridan's "regimentational" approach in comparison to the modern "formalizational" approach is that an informal system can never be as exact as a formal one, given all the possible ambiguities and arbitrariness of an informal approach. By contrast, in the formal approach, the rules of interpretation in a formal semantics and the manipulations with formulae in a formal syntax are fixed by the highest standards of mathematical exactitude, which can never be matched by any sort of informal approach. Therefore, it seems that Buridan's approach suffers from an inherent *inexactitude* that can be overcome only by the formalizational approach.

Again, Buridan's approach renders the construction of logical theory in a fundamental sense *unfinishable*. Given the immense variety and variability of the syntactical forms of a natural language, a logical theory based on its regimentation will never cover *all* syntactically possible constructions in a natural language. By contrast, since in an artificial language we have an explicit and effective set of construction rules, we can formulate logical laws that apply to all possible well-formed formulae of that language without having to worry about possible formulae that may not be covered by these laws.

Furthermore, Buridan's approach seems to be plagued by what may be termed its *linguistic provincialism*. If logical rules and distinctions are formulated in terms of the regimentation of the existing syntactical devices of a particular natural language, then, given the obvious syntactical diversity of natural languages, this approach seems to threaten the universality of logical theory. Indeed, following the lead of the syntax of a particular natural language may even provide "false clues" concerning what we may mistakenly believe to be the universal conceptual structure of *Mentalese*. By contrast, the formalizational approach provides equal access for speakers of all languages to *the same* "conceptual notation", which directly reflects the structure of the common mental language of all human beings endorsed by Buridan. So, apparently, even Buridan's logic would be much better off if it were also couched in an artificial, formal language.

Finally, if we use the natural language embodying our logic in our *reflections* on *the same* natural language, then we are obviously running the risk of Liar-type paradoxes, which are bound to emerge under the resulting conditions



of *semantic closure*, first diagnosed as such by Alfred Tarski.<sup>10</sup> By contrast, an artificial language embodying our logical theory can serve as the *object language* of the considerations concerning the syntax and semantics of this language which are to be carried out in a distinct *meta-language*. In this way we avoid the risk of paradox, since keeping the object language apart from the meta-language eliminates semantic closure.

Perhaps, these would be the most obvious reactions against Buridan's "regimentational" approach coming from someone comparing it to the modern "formalizational" approach. Nevertheless, these considerations may not be sufficient to establish once and for all the "absolute superiority" of the modern approach over Buridan's. For if we take a closer look at the modern practice, we can see that it is not much better off concerning these issues.

It must be conceded at the beginning that the mathematical exactitude of a formal logical system is unmatched by any "natural" logic (i.e. a logical system based on a certain regimentation of reasoning in some natural language). But the exactitude in question concerns only the formal system in and of itself. Concerning the formal system, we may have absolutely rigorous proofs of consistency or inconsistency, completeness or incompleteness, etc., which we may never have concerning an "unfinishable" system of "natural" logic. However, as soon as we use a formal logical system to represent and evaluate natural language reasoning, the uncertainties and ambiguities of interpretation return with a vengeance, as anyone who has ever tried to impart "formalization skills" in a symbolic logic class can testify. "Formalization" is the largely intuitive process of translating natural language sentences to formulae of a formal language based on the linguistic competence of the speakers of the natural language in question and their understanding of the import of the symbols of the formal language. Therefore, this process involves just as much inexactitude, uncertainty and ambiguity as does working with "unregimented" natural language expressions in general.

This difficulty can be overcome by constructing a formal syntax for an interesting portion of a natural language, in the vein of the approach of Richard Montague and Imre Ruzsa,<sup>11</sup> which can then serve as the basis for an exact and effective translation procedure into the artificial language of a formal logical system. In this way, having a formally constructed (not to say, "regimented") part of a natural language at our disposal, the problem of the inexactitude of

<sup>10</sup> Cf. (Tarski 1944). The gist of the idea of semantic closure is that a language that contains its own truth-predicates and has the means of referring to its own sentences is semantically closed, which is quite obviously the case with natural languages. According to Tarski, in a semantically closed language, Liar-type paradoxes ("This sentence is false" – is this true or false?) are bound to arise. For a more recent, generalized version of Tarski's argument, see (Priest 1984).

<sup>11</sup> Cf. (Montague 1973), (Montague 1974), (Ruzsa 1989).

the otherwise merely intuitive formalization process can certainly be avoided. However, given that the formal syntax can only cover a sufficiently interesting, yet manageable, part of a natural language, this approach brings out most clearly the inherently “unfinished” character of the enterprise as far as the representation of *all possible forms of natural language reasoning* is concerned. Thus, the formalizational approach can overcome the problems of inexactitude only by carving out a manageable part of natural language reasoning, thereby making explicit the “unfinished” character of the enterprise. Buridan’s regimentational approach, in comparison, simply acknowledges from the start that it can only explicate and regulate certain manageable types of natural language reasoning, and it does this with the requisite degree of exactitude, yet without introducing the explicit, full-fledged formal syntax of an artificial language that would allegedly universally reflect the structure of Mentalese underlying all natural linguistic structures.

Since the process of formalization as it is commonly practiced is based on the linguistic competence of the speakers of particular natural languages, it involves just as much “linguistic provincialism” as does the regimentational approach. Actually, it is quite instructive to observe the differences between different Montague-style approaches to formalization motivated by different languages, especially if they are also motivated by certain logical considerations that are “most natural” in those languages.

But we can also say that the syntax of standard predicate logic as we know it was also motivated by some fairly “provincial” linguistic considerations, namely, considerations concerning the language of mathematics, rather than any actual natural language. This is probably the historical reason for the notorious “mismatch” between the syntax of predicate logic on one hand, and the syntax of various natural languages on the other, which otherwise agree among themselves in those of their syntactic features that predicate logic systematically fails to match. Consider again sentences (5)-(8) listed above:

- (5) *Omnis homo qui est albus currit*  $\leftrightarrow$  (5') *Omnis homo albus currit*
- (6) *Every man who is white runs*  $\leftrightarrow$  (6') *Every white man runs*
- (7) *Omnis homo currit qui est albus*  $\leftrightarrow$  (7') *Omnis homo currit et ille est albus*
- (8) *Every man, who is white, runs*  $\leftrightarrow$  (8') *Every man runs, and he is white*

In modern predicate logic, these sentences have to be represented in terms of the basic vocabulary of the formal language of this logic. In that language, besides the logical constants (which Buridan would recognize as syncategorematic terms, such as negation, conjunction, conditional, etc.), we have primitive symbols referring to individuals, namely, individual names (intuitively corresponding

to proper nouns) as well as variables (roughly corresponding to pronouns),<sup>12</sup> and predicates (corresponding to common terms). All complex expressions are built up from these primitive symbols by means of an explicit set of rules that effectively determine which strings of these symbols are to be regarded as well-formed. Frege's original rationale for this type of construction was that he regarded all common terms as functional expressions: on this conception, a common term, such as 'man', denotes a function from individuals to the two truth-values, the True and the False. Thus, the term itself is essentially predicative; it needs to be completed with a referring expression picking out an individual to yield a complete sentence that denotes one of these truth values. Therefore, since for Frege *all* common terms denote functions of this sort, *all common terms are essentially predicative*. Accordingly, in the sentences above, even their grammatical subject terms need to be construed as predicates of individuals, which are picked out by variables bound by the quantifier word 'every' or 'omnis'. It is for this reason that universal sentences in this logic are to be represented as universally quantified conditionals. Since the subject terms of these sentences are not regarded as having the function of restricting the range of individuals to be considered in determining whether the sentence is true, these sentences will have to be interpreted as concerning all individuals in the universe, stating of them all conditionally that *if* they fall under the subject, *then* they also fall under the predicate.

Thus, (6) and (8), and the corresponding Latin sentences as well, would on this approach be parsed as expressing the same as

(6'') For everything (it holds that) if it is a man and it is white, then it is running

(8'') For everything (it holds that) if it is a man, then it is white and it is running

In other words, using the variable  $x$  in place of the pronoun,

(6''') For every  $x$ , if  $x$  is a man and  $x$  is white, then  $x$  is running

(8''') For every  $x$ , if  $x$  is a man, then  $x$  is white and  $x$  is running

And these, using the symbols of predicate logic, directly yield the matching formulae:

(6''''')  $(\forall x)[(Mx \ \& \ Wx) \supset Rx]$

(8''''')  $(\forall x)[Mx \supset (Wx \ \& \ Rx)]$

<sup>12</sup> The problems of representing anaphoric pronouns with bound variables of quantification theory generated a whole new field of research in the eighties, primarily inspired by Peter Geach's reflections on "donkey-sentences", coming from medieval logic, and especially from Buridan. For a summary account of those developments and their comparison to Buridan's ideas, see (Klima 1988).

However, given Buridan's radically different conception of the semantic function of common terms, he would provide a very different parsing for (6) and (8) (or rather for (5) and (7)). For on his conception, common terms have the function of signifying several individuals indifferently (as opposed to singular terms that would signify one individual as distinct from any other), and correspondingly they *supposit*, i.e., stand for (some of) these individuals in the context of a proposition in which the term is actually used for this purpose. Therefore, on Buridan's reading, (6) and (8) (or rather (5) and (7)) do not make a conditional claim about all individuals in the universe, but rather a categorical claim about a restricted range of individuals, namely, those marked out by the subject term, i.e., the *supposita* of the subject.

In fact, as anyone checking her own linguistic intuitions in English can testify, Buridan's analysis, coming from a "provincial" natural language, namely, Latin, matches much better the intuitions of speakers of another "provincial" natural language, namely, English. For English speakers would also find it "more natural" to understand the corresponding sentences as being categorical claims about a restricted range of individuals, rather than conditional claims about absolutely everything. To be sure, further reflection on the implications of this sort of analysis may further influence one's judgment on what "the correct" analysis of these sentences ought to be, but at least it should be clear that the Fregean analysis is definitely *not the only possible* or even the "most natural" one.

Thus, the Fregean analysis, being only one possible theoretical option, turns out to be just as provincial as Buridan's approach based on a particular natural language. Nevertheless, one may still object that at least for the Fregean analysis we have a working formal system with all the advantages of mathematical exactitude going for it, whereas we have nothing comparable for Buridan's approach. But this is simply not true.

#### BURIDAN'S LOGIC AS A LOGIC OF RESTRICTED QUANTIFICATION

As I have argued in several earlier papers, a simple, conservative extension of predicate logic can go a long way toward capturing in an exact form much of medieval logic in general and Buridan's logic in particular. Once we enhance the language of standard predicate logic with *restricted variables*, and provide the appropriate formal interpretation for their semantic evaluation in a formal semantic system, the resulting system at once becomes capable of capturing an enormous amount of traditional logic, and especially Buridan's version of it. We do not have to go into the technical details of constructing that formal system<sup>13</sup>

<sup>13</sup> For the technically-minded reader, a semantic system of this sort is available in (Klima 2001).

to explain its basic intuitive idea and its important philosophical implications concerning the relationships between this “enhanced predicate logic”, classical predicate logic, and Buridan’s informal logic.

The “basic intuitive idea” can be articulated in the following principles of construction:

- (1) Restricted variables function as variables in classical predicate logic, i.e., they are quantifiable terms that fill in the argument places of predicate letters.
- (2) Restricted variables have the general form of ‘ $v.Av$ ’, where  $v$  is what is referred to as the operator variable of the restricted variable, and ‘ $Av$ ’ as the matrix of the restricted variable, which is a well-formed formula open in  $v$  (i.e., having at least one occurrence of  $v$  that is not bound by a quantifier). The operator variable may itself be a restricted variable, in which case we can refer to it as a “nested” restricted variable (a restricted variable “nested” in another); other restricted variables occurring in the matrix of a restricted variable are spoken of as “embedded” in that restricted variable.
- (3) Restricted variables pick their values in a value-assignment *not* from the entire domain of interpretation (“universe of discourse”), but from the extension of their matrix, i.e., from the set of individuals of which the matrix is true (under a certain value-assignment of variables).
- (4) If the extension of the matrix of a restricted variable is empty, then the restricted variable has no value (which in the formal system can be represented by assigning an arbitrary value to it, outside the domain of interpretation, a so-called “zero-entity”, a technical device I owe to Imre Ruzsa). When a restricted variable has no value (i.e., technically, its value is outside the domain of discourse), then its value cannot fall within the extension of any predicate, i.e., all simple affirmative predications containing this variable in the argument of a predicate letter will come out as false.

Having these “principles of construction” in place, we can obtain a system that (i) reflects more faithfully the syntax and semantics of natural languages than standard predicate logic,<sup>14</sup> (ii) naturally extends itself to a generalized quantification theory, (iii) it not only matches, but surpasses standard predicate logic in expressive power, and (iv) provides an analysis of categorical propositions perfectly in tune with Aristotelian logic, validating all relations of the traditional *Square of Opposition* and the traditionally valid syllogistic forms.

Let us now take these four points in turn, and see exactly how the system constructed in accordance with (1)-(4) can obtain these results.

(i) Predicate logic formulae using unrestricted quantification exhibit a compositional structure involving propositional connectives that are nowhere

<sup>14</sup> For a precise characterization of the notion “faithfulness” involved in this intuitive claim, see (Klima 1988).

to be found in the corresponding natural language sentences (be they English, Latin or even Hungarian, etc.). If we take a look, e.g., at (6'''''')-(6) above, the structural mismatch is obvious. But the same sort of mismatch becomes even more striking if we change the quantifier from universal to particular (or "existential"), which requires that the main conditional be replaced with a conjunction in the resulting formula, whereas no such change is apparent in the syntax of the corresponding natural language sentence. Indeed, the variation of the natural language determiner does not require any change at all in the rest of the sentence, whereas changing the corresponding quantifier always requires a change in the propositional connectives of the formula following it, if a corresponding formula can be produced at all.

Therefore, there is no single propositional connective that could fill the place of the question mark in the following semi-formal schemata:

For every $x$	}	
For some $x$	}	
For the $x$	}	$Fx ? Gx$
For most $x$	}	
For five $x$	}	

so that we would get correct representations of the following sentence-schemata, which obviously exhibit a uniform structure (just as would the corresponding Latin, etc.):

(1) Every	}	
(2) Some	}	
(3) The	}	$F('s) \text{ is/are } G('s)$
(4) Most	}	
(5) Five	}	

Among these schemata, (1) and (2) can be represented in predicate logic only with formulae involving different propositional connectives, (3) and (5) demand complex formulae to provide their correct truth-conditions (such as the Russellian formula:  $(\exists y)\{[Fy \ \& \ (\forall x)(Fx \rightarrow x = y)] \ \& \ Gy\}$  for (3)), and for (4) there is demonstrably no quantificational formula that would provide its correct truth-conditions.<sup>15</sup>

(ii) By contrast, in the system of predicate logic enhanced with restricted variables (as well as with the requisite set of quantifiers), the following formula schema provides an intuitive formalization of (1)-(5):  $(Qx.Fx)(Gx)$ . This states that  $Q x$  that is an  $F$  is a  $G$ , or in the plural form, that  $Q x$ 's that are  $F$ 's are  $G$ 's, where  $Q$  stands for any of the appropriate determiners or "quantifier words"

<sup>15</sup> For the proof, if "most" is understood as "more than half of the", see (Barwise and Cooper 1981, 214-215.)

of English (and *mutatis mutandis* the same goes for any other natural language). This immediately establishes the claim that this system naturally extends itself to a generalized quantification theory.<sup>16</sup>

(iii) People who argue for the superiority of modern predicate logic over “traditional”, Aristotelian logic often refer to (various versions of) De Morgan’s famous example as proof that the Aristotelian analysis of categorical propositions, and correspondingly Aristotelian syllogistic, is incapable of handling reasoning involving relational terms. Intuitively, the following looks like a valid inference: ‘Every man is an animal; therefore, every man’s head is an animal’s head’. However, there is no way of parsing this inference along traditional lines so it would fit into a valid Aristotelian syllogistic form.

Medieval logicians, taking their cue from Aristotle’s *Prior Analytics*, treated such inferences under the heading *de syllogismis ex obliquis*, i.e., “on syllogisms involving oblique terms”, which is to say, terms in cases other than the nominative case, such as the genitive “man’s” in the conclusion of De Morgan’s example.<sup>17</sup> To be sure, “standard syllogistic” treats the terms of a syllogism as unbreakable units (just as propositional logic treats atomic sentences as such units), although it allows complex terms as substituends of such units. Therefore, when the validity of an inference turns on the conceptual connections between parts of such complex terms, “standard syllogistic” is indeed inapplicable (just as uniform quantification theory, involving only monadic predicates, is unable to handle inferences with multiply quantified sentences.) So, to account for such inferences, Buridan and others distinguished between the *terms of the syllogism* and the *terms of the propositions*, where the *terms of the syllogism* (in particular, the middle term) can be parts of the *terms of the propositions*, and provided further syllogistic rules in terms of this distinction, referring to the intrinsic complexity of the terms of the propositions involved.

Correspondingly, the predicate logic with restricted variables inspired by Buridan provides a compositional semantics for formulae that represent the internal structure of propositions with complex terms. Therefore, this logic has no more difficulty in handling such inferences than standard predicate logic does. There are, however, some important and instructive differences between the two.

<sup>16</sup> For good surveys of the booming research on generalized quantifiers in the mid-eighties, see (Van Benthem and Ter Meulen 1985), and (van Benthem 1986). For a recent survey of later developments see (Westerståhl 2005).

<sup>17</sup> For Buridan’s treatment, see SI) 5. 8.

In standard predicate logic, the De Morgan-example can be reconstructed as follows:

$$(\forall x)(Mx \rightarrow Ax)$$

For every  $x$ , if  $x$  is a man, then  $x$  is an animal

---


$$(\forall x)(\forall y)[(Mx \ \& \ Dxy) \rightarrow (Ax \ \& \ Dxy)]^{18}$$

For every  $x$  and every  $y$ , if  $x$  is a man and  $y$  is the head of  $x$ , then  $x$  is an animal and  $y$  is the head of  $x$ .

Using restricted variables, the same example can be reconstructed in the following way:

$$(\forall x.Mx)(\exists y.Ay)(x = y.)^{19}$$

Every ( $x$  that is a) man is (identical with) some ( $y$  that is an) animal

---


$$(\forall x.( \exists y.My)(Dxy.))(\exists u.( \exists v.Av)(Duv.))(x = u.)$$

Every ( $x$  that is a) head of some ( $y$  that is a) man is (identical with) a ( $u$  that is a) head of some ( $v$  that is an) animal<sup>20</sup>

One important difference between these two reconstructions is that if we drop the parenthetical phrases in the semi-formal sentences that are simply transcribed into the formulae with restricted variables, then we get perfectly good English sentences, which cannot be done with the semi-formal sentences transcribed into the standard formulae. This quite clearly indicates the close match between the syntax of the natural language sentences and the formulae with restricted variables.

Another important difference is that while the standard formulae are true if there are no men or they have no heads, those with restricted variables in those circumstances would be false. Therefore, according to the formalization with restricted variables, the inference is not formally valid, unless there is a further premise to guarantee that if there are men, then there are men's heads. Actually, this is how it should be. After all, even if it is actually true, it is not a *logical truth* (i.e., a truth based on the meaning of logical connectives) that if there are men, then they have heads. Therefore, the formulation with restricted variables provides an even better analysis of the natural language sentences, in the sense that it better reflects our semantic intuitions as to what is and what is not implied by the sentences in question.

<sup>18</sup> I am providing here the "stronger", but "more intuitive" formalization of this sentence. Cf. (Merrill 1977).

<sup>19</sup> To simplify formulae with restricted variables, the matrix of a restricted variable may be omitted after its first occurrence.

<sup>20</sup> For a similar analysis with the same results, see (Orenstein 2000). For a detailed discussion of the neat syntactical match between restricted quantification and natural language sentences, see (Klima 1988).



Thus, we have to conclude that the “Buridan-inspired” predicate logic with restricted variables, besides covering more than standard predicate logic does as far as non-standard quantifiers are concerned, can handle what standard predicate logic can, indeed, while sticking more faithfully to the syntactic construction of natural languages and reflecting better our semantic intuitions concerning reasoning in natural languages.

(iv) What accounts for the difference between the judgments of the two different formalizations concerning the validity of De Morgan’s example is their difference in attributing vs. denying existential import to universal affirmative propositions. The reason why De Morgan’s example at first appears to be intuitively valid is that we tend to tacitly assume the non-logical truth that if there are men, then they have heads too. However, a formally valid inference has to yield truth from truth with any terms, which is actually not obvious with De Morgan’s example. Consider the following, analogous example: ‘Every man is an animal; therefore, every man’s hat is an animal’s hat’. Suppose there are men, but no man has a hat, which is certainly possible. In that case it is obviously true that every man is an animal, but is it true that every man’s hat is an animal’s hat? Or take the following, perhaps even more obvious example: ‘Every horse is an animal; therefore, every horse’s wing is an animal’s wing’. Knowing that there are no winged horses, and hence no horse’s wings, we would naturally tend to reject the conclusion. To be sure, one may still understand this conclusion conditionally, as saying that *if* something is a horse’s wing, *then* it is an animal’s wing, but that conditional reading would lose precisely the matter-of-fact character of the original categorical claim.

Indeed, other examples can bolster our intuition that even if universal affirmatives may occasionally have the force of a conditional, hypothetical claim, especially when they are supposed to express a law-like statement; nevertheless, it is simply wrong to assume that they *always* have to be interpreted this way. Consider for example the case of Mary boasting to her friends that every boy kissed her at the party yesterday. If her friends later find out that there were no boys at the party, then they will certainly take her for a liar, rather than accept her claim as being “vacuously” true on account of her universal claim expressing a universally quantified conditional with a false antecedent. Such and similar examples could be multiplied *ad nauseam*. What is important, though, is the fact that we do have the intuitive distinction between the categorical and hypothetical readings of universal affirmatives; therefore, a logic that can acknowledge both of these readings is certainly preferable to one that can only handle one of them. Since predicate logic with restricted variables is a conservative extension of standard predicate logic in the sense that all formulae of the standard logic are formulae of the logic enhanced with restricted variables, the latter is of course capable of representing whatever the former can, but not *vice versa*.

## CONCLUSION: BURIDAN'S "NATURAL LOGIC" VS. ITS RECONSTRUCTION

Nevertheless, although this has to be the end of this lecture, this is far from being the end of the story of comparing medieval and modern logic. The foregoing could serve merely to illustrate that heeding medieval logicians' *regimentation of natural language*, we may be able to come up with some *more natural formalization* in a *formal language*. But, as I argue in detail in my monograph on John Buridan (Klima 2009), where this lecture comes from, Buridan would still not be quite happy with this formal reconstruction of his logic. And the reason would not be its formalism (after all, Buridan also uses some symbolism time and again), but rather its restricted applicability in other areas, where Buridan's logic still has important lessons to teach us. In particular, even if quantification theory with restricted variables can easily be extended to cover a great deal of Buridan's modal and temporal logic, it cannot quite properly handle Buridan's treatment of intensional contexts generated by words signifying our mental acts. Moreover, Buridan would not be quite happy with restricted variables representing his common terms, since for him there are *simple* common terms, say, *F*, the semantic properties of which are different from a *complex term*, such as 'an *x* that is an *F*'. In fact, in various contexts, Buridan would sharply distinguish between the logical import of the two. Finally, and even more importantly, Buridan would reject both the Quinean idea of ontological commitment usually associated with quantification theory and the global distinction between object language and meta-language, built into the very construction of this theory. As I argue in my book, this double rejection allows Buridan to work out a third alternative "between" a Quinean and a Meinongian approach to ontological commitment, as well as a viable logical theory for semantically closed natural languages, avoiding Liar-type paradoxes. But this much may be just enough by way of "a shameless plug" to finish this lecture.

## REFERENCES

Aristotle, *On Interpretation*

SD = Johannes Buridanus, *Summulae de Dialectica*. Translation used: Klima, G., John Buridan: *Summulae de Dialectica*, an annotated translation with a philosophical introduction; New Haven, Yale University Press, 2001.

Barwise, J. and Cooper, R., 1981, Generalized Quantifiers and Natural Language. *Linguistics and Philosophy* 4(1981). 159—219.

Klima, G., 1988, Essay III.: General Terms in their Referring Function. In id., *Ars Artium: Essays in Philosophical Semantics, Medieval and Modern*, Budapest, Institute of Philosophy of the Hungarian Academy of Sciences. 44—84.

- Klima, G., 1991, Latin as a Formal Language: Outlines of a Buridanian Semantics. *Cahiers de l'Institut du Moyen-Âge Grec et Latin* 61. 78—106.
- Klima, G., 2001, Existence and Reference in Medieval Logic. In A. Hieke and E. Morscher (eds.), *New Essays in Free Logic*. Dordrecht, Kluwer Academic Publishers. 197—226.
- Klima, G., 2009, *John Buridan*. Oxford, Oxford University Press.
- Merrill, D., 1977, On De Morgan's Argument. *Notre Dame Journal of Formal Logic* 18. 133—139.
- Montague, R., 1973, The Proper Treatment of Quantification in Ordinary English. In J. Hintikka, J. Moravcsik and P. Suppes (eds.), *Approaches to Natural Language*. Dordrecht, Reidel. 247—270.
- Montague, R., 1974, English as a Formal Language. In R. Thomason (ed.), *Formal Philosophy*. New Haven-London, Yale University Press.
- Orenstein, A., 2000, The Logical Form of Categorical Sentences, *Australasian Journal of Philosophy* 78, 517—533.
- Priest, G., 1984, Semantic Closure. *Studia Logica* 43. 117—129.
- Ruzsa, Imre 1989, *Logical Syntax and Semantics* vol. II. (in Hungarian) Budapest, Akadémiai Kiadó.
- Tarski, A., 1944, The Semantic Conception of Truth. *Philosophy and Phenomenological Research* 4. 342—375.
- van Benthem, Johan and Alice ter Meulen (eds.), 1985, *Generalized Quantifiers in Natural Language*. Dordrecht: Foris Publications.
- van Benthem, Johan, 1986, *Essays in Logical Semantics*. Dordrecht, Reidel.
- Westerståhl, D., 2005, Generalized Quantifiers. In *The Stanford Encyclopedia of Philosophy* (Winter 2005 Edition), ed. E. N. Zalta, URL = <http://plato.stanford.edu/archives/win2005/entries/generalized-quantifiers/>.

## Frege's Definition of Number: No Ontological Agenda?

**Abstract.** Joan Weiner (2007) has argued that Frege's definitions of numbers constitute linguistic stipulations that carry no ontological commitment: they don't present numbers as pre-existing objects. This paper offers a critical discussion of this view, showing that it is vitiated by serious exegetical errors and that it saddles Frege's project with insuperable substantive difficulties. It is first demonstrated that Weiner misrepresents the Fregean notions of so-called *Foundations*-content, and of sense, reference, and truth. The discussion then focuses on the role of definitions in Frege's work, demonstrating that they cannot be understood as mere linguistic stipulations, since they have an ontological aim. The paper concludes with stressing both the epistemological and the ontological aspects of Frege's project, and their crucial interdependence.

### 1 THE PROBLEM

It is indisputable that Frege's logicist project, including the development of his logical calculus, had an epistemological aim, namely to prove the *a priori* and analytic status of the arithmetical truths, and thus to prove that they are deducible from the laws of logic. More problematic, and a subject of recent debate, is the question concerning the status of definitions within this project.<sup>1</sup> Frege dismisses previous attempts at the definition of number, and replaces them with new definitions. In addition, in several passages he describes definitions as arbitrary conventions. So it seems as if Frege is not interested in capturing with his definitions the pre-existing meanings of arithmetical symbols, but in stipulating new ones. But how can this revisionary project be brought into harmony with the epistemological aim, i.e. how can arbitrary

<sup>1</sup> Some key contributions to this debate are (Benacerraf 1981), (Weiner 1984), (Picardi 1988), (Kemp 1996). For more details, see the recent overview in Shieh (2008).

definitions contribute to proving the logical status of *the* truths of arithmetic, i.e. of antecedently existing truths?

In the most recent contribution to this debate Joan Weiner (2007) offers a radically new solution: Frege was a more thorough revisionist than the dilemma above presents him. His revisionism affected not only his conception of definition, but also of sense, reference and truth. Prior to his work, numerals did not have a determinate sense and reference, and arithmetical statements were not strictly speaking true. According to Weiner, Frege did not believe that concept-script systematisation is unveiling the true nature of numbers and the true referents of numerals, but only that it introduces stricter semantic and inferential constraints of precision stipulating the sense and reference of numerals and arithmetical statements for the first time. Thus talk about numbers as objects and strict arithmetical truth is only possible as a system-internal discourse, and concept-script systematisation is a normative linguistic precisification serving an epistemological aim, with no ontological and semantical discoveries about pre-systematic arithmetic and its language. In particular, definitions carry no content-preserving and ontological commitment.

## 2 WEINER'S ARGUMENT

Weiner offers a wealth of substantive and exegetical considerations in favour of her view, focusing most explicitly on the role of definitions within Frege's work, especially in the *Foundations of Arithmetic*. She investigates what requirements a definition (of number, numerals etc.) must satisfy in order to qualify as adequate or faithful to prove the truths of arithmetic from primitive truths (Weiner 2007, 683). One such obvious requirement seems to be the following:

*The obvious requirement:* A definition of an expression must pick out the object to which the expression already refers or applies (ibid. 680).

Weiner denies there is any evidence in Frege's writings for this requirement. Definitions are not preserving the putative pre-systematic reference of numerals. Still, they must be faithful to pre-systematic arithmetic in some sense, since systematisation is not meant to transform arithmetic into some 'new and foreign science' (ibid. 687). Her explanation is as follows: 'Faithful definitions must be definitions on which those sentences that we take to express truths of arithmetic come out true and on which those series of sentences that we take to express correct inferences turn out to be enthymematic versions of gapless proofs in the logical system' (ibid. 690, 790). In other words, what systematisation preserves is truth-related and inference-related content. For example, regarding truth-related content a definition of '0' and '1' is unacceptable, if it presents as true a sentence

which in pre-systematic arithmetic is taken to express a falsehood, namely ' $0=1$ '. Thus faithful definitions must cover for what are taken to be the well known properties of numbers.<sup>2</sup> Regarding inference-related content, faithful definitions must preserve inferences that we take to be valid, for example 'If Venus has zero moons and the Earth one, then given that  $0 < 1$ , the Earth has more moons than Venus' (ibid. 686). Thus faithful definitions must cover for all applications of number, including those in applied arithmetic.<sup>3</sup>

Frege is therefore not concerned with preservation of reference, and not even simply with preservation of putative truths, rather of what Weiner calls '*Foundations-content*'. This is 'some sort of content connected with inferences' (ibid. 692). *Foundations-content* partly points back to the judgeable content of *Begriffsschrift*, which was defined by Frege as content that has only 'significance for the inferential sequence' (1879, x.). But *Foundations-content* also partly anticipates the later notion of sense, i.e. *Sinn* (Weiner 2007, 689-1), for two reasons. First, the judgeable content of a term, she claims, is not its referent (ibid. 690, fn. 17), just as much as sense is not reference. Second, a term hitherto considered non-empty will not cease to have *Foundations-content* if we discover it is empty, for the discovery will lead to a re-evaluation of our pre-systematic beliefs and inferences, a re-evaluation still involving the term itself (ibid. 690). Equally, a fictional term like 'Hamlet' has *Foundations-content*, since there are speakers who think it enables them to express truths and correct inferences (ibid. 691). Hence, it is not required for a term to have a referent in order to have *Foundations-content*, and this brings *Foundations-content* in the vicinity of *Sinn*.

Thus, what concept-script systematisation achieves is preservation of *Foundations-content*. However, this should not be understood in the trivial sense of 'preservation', as if something outside of the system is identical with something in the system. As it transpires from Weiner's argument, preservation of *Foundations-content* means rather something like 'normative transformation of pre-system content into systematic content'. As quoted above, systematisation involves the process of proving *within the logical system* the truth of the pre-systematic sentences taken to express truths as well as the correctness of pre-system inferences taken to be correct. But proving 'in the logical system' is a highly normative process, guided essentially by *two precision requirements* that distinguish sharply the system from the pre-system: the gapless proof requirement, i.e. all proofs are absolutely gapless, and the sharpness requirement, i.e. genuine concepts must have sharp boundaries.<sup>4</sup> Essentially, this 'preservation' is to be understood as a creative process of precisification of pre-systematic language,

<sup>2</sup> (Weiner 2007, 688.) Cf. (Frege 1879, §70).

<sup>3</sup> (Weiner 2007, 689.) Cf. (Frege 1879, §19).

<sup>4</sup> See (Weiner 2007, 701), (Frege 1884, §62, §74), (Frege 1903, §56).

which is indeterminate and vacillating (ibid. 697), as himself Frege seems to suggest about expressions quite generally (see (Frege 1906a, 302–3)). 'Frege's task is to replace imprecise pre-systematic sentences with precise systematic sentences [of arithmetic]' (Weiner 2007, 710). This suffices to make sense of Frege's epistemological aim, the core of his logicist project.

This view has some intriguing implications and corollaries, the most important of which shall be briefly summarised. More details evidence will be presented during the discussion further below.

- (a) *Foundations*-content is close to sense, but not identical with it. To have a determinate sense, an expression must have a definition satisfying the precision requirements. But pre-systematic expressions don't have such a definition, hence they don't have a sense and, by extension, no reference.<sup>5</sup> Sense and reference (*Sinn* and *Bedeutung*) are therefore only system-internal features of expressions.<sup>6</sup> There is no evidence to the contrary in Frege's writings. In particular, Frege never says that terms of pre-systematic language have a reference or require one in order to have a use (ibid. 706f.).<sup>7</sup> The absurdity of this view is merely apparent, for fixing the sense and reference of a term is only the ideal end of a science, once it comes to fruition in a system (ibid. 709f.).<sup>8</sup>
- (b) If pre-systematic terms don't have a determinate reference, then given compositionality, pre-systematic sentences don't have a determinate reference either, i.e. a truth-value. We only 'take them to express truths' (ibid. 690f.). This does not mean there is nothing 'right' about them (ibid. 710), but only that their rightness does not satisfy the constraints imposed by systematisation. We must distinguish between different notions of truth, as Frege does, i.e. *pre-systematic truth* and *strict truth* (ibid. 709f.).<sup>9</sup> Pre-systematic truth is one of the aforementioned faithfulness requirements a definition (of number etc.) must satisfy.
- (c) Since pre-systematic arithmetical expressions do not have a determinate reference, ordinary arithmetical predicates like 'is a number' do not have a reference either. Hence, the concept of number is not already fixed prior to Frege's definitions (ibid. 696). Quite the opposite: in *Foundations* (§100) Frege stresses the arbitrary, stipulative character of definitions (ibid. 695ff.), and he does so again in the important posthumous

<sup>5</sup> The alternative of having an indeterminate sense (and reference) is excluded for Frege, since there is no such thing for him. See for instance (1903, §56).

<sup>6</sup> This claim has been advanced before. See e.g. (Stekeler-Weithofer 1986, 8.,10.).

<sup>7</sup> Weiner's argument seems to come close here to a Wittgensteinian theory of meaning as use, although Wittgenstein is not mentioned.

<sup>8</sup> See also (Frege 1914, 242).

<sup>9</sup> 'Pre-systematic truth' is not Weiner's term, but my terminological correlate to her label 'regarding a sentence as true' (ibid. 706, 709).

text “Logic in Mathematics” (Frege 1914). Here he claims that a determination of sense is either decompositional, in which case it is a self-evident axiom, or it is a mere stipulation (‘constructive definition’). Since Frege does not seem to present his *Foundations* definitions as self-evident (1884, §69), they must be stipulations, stipulations which precisify *Foundations*-content and thus transform arithmetic into a system of science.

- (d) A cursory reading of Frege’s writings might induce one to assume that he thinks numbers are pre-existing, language-independent objects whose nature his definitions aim to capture. Call this *the ontological thesis*. Weiner rejects this thesis.<sup>10</sup> On her view Frege makes no claim that numbers existed prior to his definitions, or else he would have to say that the definitions are (or articulate) discoveries about pre-existing objects. But they are linguistic stipulations, not ontological discoveries. Frege does not claim that it is part of the nature of numbers to be extensions, but is interested only in the linguistic question ‘Are the assertions we make about extensions assertions we can make about numbers?’, which he answers by means of a linguistic principle par excellence, the context principle (Weiner 2007, 698f.). As Frege writes: ‘I attach no decisive importance to bringing the extensions of concepts into the matter at all’ (1884, §107).

Weiner’s interpretation is certainly intriguing and original. Nevertheless, it is vitiated by serious exegetical errors, and it saddles Frege’s theory of numbers with insuperable substantive difficulties. I will first show that Weiner misrepresents so-called *Foundations*-content, sense and reference, and the notion of truth in Frege’s work (sections 3-5). Then I will focus on the role of Fregean definitions, demonstrating that they have, *pace* Weiner, an ontological point, and that they are not mere stipulations. The paper concludes with stressing both the epistemological and the ontological aspects of Frege’s project, and their crucial interdependence.

### 3 ‘FOUNDATIONS-CONTENT’?

We can start with the notion of *Foundations*-content, on which Weiner bases her rejection of the obvious requirement. Is there really a notion of content in the *Foundations* closely related, although not identical to sense? There is no decisive evidence. Frege uses the term loosely. It may mean various things such

<sup>10</sup> Weiner has defended this anti-ontological stance in previous work. See for instance (Weiner 1990, ch. 5.)



as 'sense of a sentence', i.e. a judgement or thought (1884, x., §3, §70, §106), 'sense of a recognition judgement' (ibid. §106, 109), 'judgeable content' (ibid. §62, §74), 'reference' (ibid. §74fn.), and also just conventional meaning and use (ibid. §60). 'Content' in the *Foundations* is simply not a technical term, which fits the prolegomenous character of the book. There is nowhere an argument specifying that an expression could have a content but lack a referent, i.e. that 'Hamlet' has *Foundations*-content. In one place Frege claims the exact opposite in fact: 'the largest proper fraction' has no content and is senseless, because no object falls under it (ibid. §74fn.). He makes a similar point about 'the square root of -1' (ibid. §97). Moreover, the predominant and philosophically significant role of 'content' in the *Foundations* is found in Frege's repeated requests to specify a content of arithmetical judgements in such a way that they turn out to express identities, and thus to secure the objecthood of numbers, given his acceptance of the principle that identity is an essential mark of objecthood (1884, §62f.).<sup>11</sup> Thus we have positive evidence that content in the *Foundations* is closer to reference than to sense.

At the very least, is *Foundations*-content not related to *Begriffsschrift* content and insofar only inferential, not referential? But there is no dichotomy here. Judgeable content is so intimately tied to reference that it affects the most basic formation rules of the notation in *Begriffsschrift*. Frege stipulates that any expression following the content stroke must have a judgeable content. That the relation between a judgeable content and its expression is one of being designated, is visible from at least two facts: that the expression of a judgeable content is a designator starting with the definite article, paradigmatically the nominalised form of a proposition ('the circumstance that there are houses', 'the violent death of Archimedes at the capture of Syracuse')<sup>12</sup>, and that the expression of a judgeable content can flank the sameness of content sign, i.e. the identity sign. Hence, the expression of a judgeable content is a name of the judgeable content and no formula in concept-script is even syntactic if such a name fails to designate anything.<sup>13</sup>

Weiner argues that since 'Phosphorus = Phosphorus' is derivable from the law of identity, while 'Hesperus = Phosphorus' is not, the two names cannot have the same *Begriffsschrift*-content (Weiner 2007, 690). But this example is uncongenial to Frege's concerns: 'Phosphorus' and 'Hesperus' are names of *unjudgeable* content, while in *Begriffsschrift* he is interested only in names of judgeable content—content that can be asserted. Hence, unlike with names of *unjudgeable* content, there is no room for distinguishing between what a name

<sup>11</sup> See also (Rumfitt 2003, 198.)

<sup>12</sup> (Frege 1879, §2-3.)

<sup>13</sup> Frege refers to the expression of a judgeable content flanking the identity sign explicitly as a name (Frege 1879, §8, *passim*).

of a judgeable content refers to and its inferential content. Therefore, once it has been established (by a synthetic judgement) that two names name the same judgeable content, 'A' and 'B' are intersubstitutable for the purposes of inference (' $\equiv$ ' functions as an identity, but also as a stipulative sign; see (Frege 1879, §8, §23)), just as much as, vacuously, 'A' and 'A'. The difference between two names of the same judgeable content is neither a referential nor an inferential difference, but only of a different mode of determination of the same inferential content, as Frege explicitly states (Frege 1879, §8). Claiming that *Foundations*-content is somehow connected to *Begriffsschrift* content, while arbitrarily revising Frege's own view of the latter, is question-begging.

#### 4 SENSE AND REFERENCE

It is equally unwarranted to treat sense and reference (*Sinn* and *Bedeutung*) as system-internal features which do not apply to ordinary expressions. There are enough passages to the contrary, for example when Frege writes that 'The moon is the reference [*Bedeutung*] of "the moon"' (1919, 255) or when he explains, in a lecture, that '[i]f we claim that the sentence "Aetna is higher than Vesuvius" is true, then the two proper names do not just have a sense [*Sinn*] [...], but also a reference [*Bedeutung*]: the real, external things that are designated'<sup>14</sup>. In "Introduction to Logic" he writes: 'If we say "Jupiter is larger than Mars", what are we talking about? About the heavenly bodies themselves, the references [*Bedeutungen*] of the proper names "Jupiter" and "Mars"' (1906b, 193). In *Grundgesetze* he points out that the notion of reference (*Bedeutung*) cannot be (genuinely) explained, since any such explanation would presuppose knowledge that some terms have a reference (Frege 1893, §30); hence reference must predate the setup of any system. Since to have a reference implies having a sense, it follows that those expressions of natural language that have reference also have a sense. Neither feature is therefore exclusively system-internal for Frege. Hence, if it is insisted that Frege's definitional project is to be described as preserving so-called *Foundations*-content, then this will involve preservation of pre-systematic content that has a referential component, and this will not be content of a wholly different kind from the content described in *Grundgesetze* split into sense and reference (see (Frege 1893, x.)).

<sup>14</sup> Carnap 2004, 150.

## 5 KINDS OF TRUTH?

Weiner claims that Frege allows for many notions of truth, including pre-systematic and strict truth. But nowhere does Frege make such a claim. He argues, in “Thoughts”, that truth is undefinable, on the basis that any attempted definition would have to analyse truth into constituent properties (of the truth bearer), of which in turn it would have to be *true* that they apply in a particular case in order to make the definition applicable (1918a, 60). This circularity suggests not only the indefinability of truth, but also its simplicity, and thus its univocity. Frege intimates in “Thoughts” additional arguments—very plausible ones—against Weiner’s claim. Thus he distinguishes sharply between ‘taking something to be true’ (*Fürwahrhalten*) and ‘proving the true’ (*Beweis des Wahren*). The former arises through psychological laws, while the latter belongs to the laws of truth, which are not the object of psychology, but only of logic (ibid. 58f.). Since Weiner characterises pre-systematic truth in terms strongly resembling *Fürwahrhalten*, e.g. ‘what we take to express truth’ or ‘what we regard as true’, it would follow, absurdly, that Frege takes truths of pre-systematic arithmetic to belong to the realm of the psychological, and pre-systematic arithmetic to psychology. Equally, if pre-systematic truth is vacillating and vague, it would seem to be a predicate coming in degrees. But Frege specifies that truth does not allow for ‘more or less’ (ibid. 60). Finally, Frege speaks on repeated occasions about the truths of arithmetic or mathematics as such (e.g. (1884, §3, §11, §14, §17, §109)). But he nowhere qualifies them as merely pre-systematic, to be distinguished from the truths arrived at within the system.

Quite generally, the truth of a thought is timeless (e.g. (1884, §77, 1918a. 74)). Hence, either a pre-systematic notion of truth is timeless as well, in which case it is not clear how this squares with its indeterminacy and vagueness, or a pre-systematic assertion does not express a thought at all. Of course, Weiner can retort that this is indeed so: a pre-systematic assertion only expresses *Foundations*-content. But then *Foundations*-content is assertable. And it is negatable, thinkable, judgeable etc. However, *Foundations*-content is not really judgeable, since judging is ‘acknowledging the truth of a thought’ (ibid. 62). Judgeable is only that to which strict truth can apply. But then *Foundations*-content is not assertable either, for to assert is to make manifest a judgement (ibid.). And if it is not judgeable, *Foundations*-content lacks the essential association with judgeable content Weiner claims it has, and thus it does not have an inferential character either, for to infer is to judge (1879-1891, 3). Pre-systematic arithmetical proofs and arguments could not be counted as valid, if we continue this line of thought. In conclusion, the distinction between

pre-systematic and strict truth, in conjunction with the notion of *Foundations*-content, leads to catastrophic consequences.<sup>15</sup>

## 6 THE OBJECTIVITY OF DEFINITIONS: THE MATERIAL MODE

Doubts about Weiner's interpretation also arise if we look more carefully at what Frege says about and what he does with his definitions in the *Foundations*. There are passages in the *Foundations* whose wording stress the stipulative character of definition, e.g. when Frege speaks about fixing ('*festsetzen*') the sense or meaning of expressions, or refers to definitions as stipulations ('*Festsetzungen*').<sup>16</sup> But his phraseology indicates an objective side to definition as well. Thus he speaks of the need to ascertain, find out ('*feststellen*'), explain the sense of an equation (ibid. x., p. 73, §62, §106), which is inaccurately translated as 'to fix/to define the sense' by Austin, or of the need to find or attend to ('*aufsuchen*') a judgeable content which can be transformed into an identity whose sides contain the new numbers (1884, §104).<sup>17</sup>

The objectivity of definitions is also manifest in Frege's tendency to adopt the material mode and define objects, as opposed to mere expressions. This is confirmed by many passages in the *Foundations* (1879, §7, §8, §9, §10, §18, §67). In two places he speaks explicitly of 'the definition of an object' (1879, §67, §74). A case in point is the very passage in §100 which Weiner takes as decisive evidence in favour of her thesis that Fregean definitions are creative linguistic stipulations. She claims that Frege is telling us that the meaning of 'the square root of -1' is not fixed prior to our definitions, but only fixed for the first time by the definitions (Weiner 2007, 695). However, let us look at the full context:

We should be equally entitled to choose as further square roots of -1 a certain quantum of electricity, a certain surface area, and so on; but then we should naturally have to use *different symbols* to signify these *different roots*. That we are able, *apparently*, to create in this way as many square roots of -1 as we please, is not so astonishing when we reflect that the meaning of the square root of -1 is not something which was already unalterably fixed before we made these choices, but is decided for the first time by and along with them' (1884, §100; my italics).

<sup>15</sup> One additional problem: if we accept the distinction, what are we to do with Weiner's own arguments, which are formulated in pre-systematic philosophical prose, and not in concept-script? Are they not valid either? Are her conclusions not strictly true? Both a 'yes' and a 'no' answer invalidate her theory.

<sup>16</sup> E.g. §7, §65, §67, §68, §75, §104, §109.

<sup>17</sup> See also §106, where Frege reports that he has established that numbers are not collections of things or properties.

Clearly, Frege is discussing here the possible definition of an object. This is why he talks about different *symbols* having to be assigned to each square root of -1, *once* each square root of -1 has been chosen. So the definitional choice is not a linguistic one. In addition, Frege's tone is verging on sarcasm here, as it is in the embedding discussion.<sup>18</sup> He is certainly not telling us that a definition entitles us, in virtue of its creative powers, to introduce as many square roots of -1 as we want, rather he is presenting his opponent's point of view ('apparently').<sup>19</sup> The latter is a sort of reformed formalist, who has accepted Frege's anti-formalist arguments (up to §100), according to which the introduction of signs alone will not bring complex numbers into existence. Therefore, the reformed formalist sets out to supplement his definition of a complex number with the assignment of a random object, to 'fix' the 'meaning' of the complex number. Frege is spelling out the absurd consequences of this approach, viz. the possibility of a plurality of such assignments.<sup>20</sup> Weiner's misunderstanding of this passage is twofold: Frege is not propounding *his* own view of definition of a linguistic *symbol* as creative stipulation, but his imaginary *opponent's* view of definition of an *object* as a random ontic assignment. There is no evidence in *Foundations* that Frege takes definitions to be arbitrary, creative definitions.

Weiner falls prey to a similar misunderstanding when she discusses an eligibility condition of primitive truths. In "On Formal Theories of Arithmetic" (1885), Frege argues that every definition must come to an end, hitting upon indefinable primitives, the original building blocks of science (*Urbausteine*), which are expressed in axioms (1885, 96). Weiner takes this to be a semantic point: the eligibility condition 'is that the *expression* of the primitive truth should include only simple, undefinable *expressions*. For these simples are the ultimate building blocks of the discipline' (Weiner 2007, 682, my italics). But this is not Frege's point. Again, Frege speaks here in the material, not the formal mode, concerned with defining the objects themselves, in this case the objects of geometry: 'It will not be possible *to define an angle* without presupposing knowledge about the straight line. Of course, what a definition is based on might itself have been defined previously' (1885, 104, my italics). The primitives terminating such a chain of definitions are not expressions, but undefinable objects 'whose

<sup>18</sup> In the same context he remarks: 'Let the Moon multiplied by itself be -1. This gives us a square root of -1 in the shape of the Moon' (1884, §100). This was hardly written with a straight face, given Frege's usual predilection for sarcasm and irony. See also his related attack against Kossack (1884, §103): 'We are given no answer at all to the question, what does  $1 + i$  really mean? Is it the idea of an apple and a pear, or the idea of toothache and gout? Not both at once, at any rate, because then  $1 + i$  would not be always identical with  $1 + i$ .'

<sup>19</sup> Dummett is another author who misunderstands this passage. See (Dummett 1991, 178f).

<sup>20</sup> And he shows a few pages on that the formalist origin of this kind of reasoning leads to misconstruing the subject matter of arithmetic as synthetic and even as synthetic a posteriori (1884, §103).

properties are expressed in the axioms' (ibid.).<sup>21</sup> It is in the material mode that Frege goes on to explain that the building blocks of arithmetic must be of purely logical nature (or else we cannot account for the universality of arithmetic), and to describe the terminological replacement of 'set' with 'concept' as not being a mere renaming, but of importance for the actual state of affairs (1885, 104f.).

## 7 THE OBJECTIVITY OF DEFINITIONS: FREGE'S PLATONIC REALISM

Another challenge to Weiner's interpretation is Frege's Platonism, which he maintains with respect to arithmetical truth, arithmetical objects and logical objects, and which is manifest in the importance existence proofs play for his definitions. Frege explains the definition of 1 by reference to the sempiternality and apriority of the truth of the propositions it helps to derive, e.g. '1 is the immediate successor of 0'; no physical occurrence, including 'subjective' ones concerning the constitution of our brains, could ever affect the truth of this theorem (1884, §77). A mere linguistic constraint of precisification cannot explain this sempiternality; the truth-value of 'Tom is alive' will continue to depend on contingent, empirical facts even if we cut the boundaries of the embedded expressions sharp.<sup>22</sup> What explains the sempiternality is the specific objectivity of the content of arithmetical propositions, the nature of arithmetical objects. This nature is grounded in the nature of logical objects, given that number-statements are ultimately about relations between logical objects (correlations of (extensions of) second-order concepts). Far from explaining logical objects as the result of linguistic constraints and stipulations, he ascribes to them ontological objectivity: they are simply there, ready to be discovered by us. Thus judgeable contents, the paradigmatic logical objects of the early work, are for Frege as objective as any mind- and language-independent object, like the sun (1879-1891, 7), although not physical.<sup>23</sup> Definitions could not play any creative role at this ontic level, and they certainly don't play it in concept-script, where a definition is merely an abbreviation: it stipulates that a simple sign is to have the same judgeable content as a more complex one (1879, §24). The existence

<sup>21</sup> We can call the expressions designating such primitives also 'primitives', but only metonymically.

<sup>22</sup> Frege's eternalist theory truth, maintained elsewhere, does not matter here, since that theory is not concerned with the justification of the truth-value of a proposition, but with the question about the proper bearer of truth.

<sup>23</sup> Elsewhere he says something similar about the relation between an object and its concept: 'To bring an object under a concept is merely to recognise a relation that already existed beforehand' (1984, 198). Also: 'Our relation to logical truths and mathematical structures is inessential to their nature and existence' (1984, 371). The beautiful passage in the preface to *Grundgesetze*, according to which the laws of logic 'are boundary stones set in eternal foundations, which our thought can overflow, but never displace' (1893, xvi.), is also relevant here.

of judgeable contents and of names of judgeable contents thus precedes any definition in concept-script.

Frege claims the same kind of objectivity for numbers in the *Foundations*, i.e. independence of the mind, language, stipulations. Corresponding passages are legion (e.g. §26, §60, 62), and it is hard to see how mere precision requirements imposed by system construction could make sense of them. Take the remarkable passage in which he compares the objectivity of number with that of the North Sea, pointing out the independence of both from our arbitrary stipulations (1884, §26).<sup>24</sup> If we were to slightly change the meaning of 'North Sea' today, whatever true content (thought) has been expressed until now by 'The North Sea is 10,000 miles in extent' would not become false. The North Sea is out there, objective, ready to be discovered by us. Presumably, then, there are correct and incorrect definitions of the North Sea, if the North Sea is independent of the definition of 'the North Sea': a correct definition will pick out precisely the North Sea. *Pace* Weiner, what this suggests is that there is an element of discovery in at least a subclass of definitions: the correct ones pick out a pre-existing object in a precise and determinate manner. This is entirely compatible with what Frege says elsewhere about numbers, namely that we discover them in the concepts (1884, §48, §58), and about the similarity between the mathematician and the geographer: neither can create 'things at will; [both] can only discover what is there and give it a name' (1884, §96).<sup>25</sup> It is unclear how Weiner's interpretation can cope with such passages. They are not mentioned in her article.

As is visible from his rejection of empiricism in logic and mathematics, the objectivity Frege claims for logical and arithmetical entities is actually stronger than that of physical entities. Numbers are abstract objects, ready to be recognised by us, but without physical properties, including spatiality and temporality. Our recognition of abstracta is itself situated in time (and presumably space), but abstracta are not, as he explains using the example of the equator. The equator has not been created in the sense that nothing positive could be said about it prior to its alleged creation (1884, §26). This example is not wholly fortunate, since the equator is a dependent abstractum (prior to the creation of the Earth there was indeed nothing positive to say about the equator), but Frege's point is on firm grounds with respect to self-subsistent abstracta like numbers: there is something positive to say about them at all times (with the appropriate linguistic usage in place). This renders the temporalisation of the truth of statements concerning numbers (e.g. "Numbers are extensions" is not

<sup>24</sup> Note, and ignore, that Frege's discussion of the arbitrariness of stipulations is confusingly embedded in a discussion against psychologism.

<sup>25</sup> Frege also compares the mathematician with the botanist who determines something objective when he determines the number and colour of a plant's petals. See also (1893, xiii.)

a true statement prior to Frege's definitions formulated in 1879'), as entailed by Weiner's argument, both objectionable and unFregean.

Frege's drive towards a Platonist ontology is also manifest in another respect. In the *Foundations* Frege does not merely provide us with a definition of, say, 0, and content himself with the fact that it satisfies the sharpness requirement. Instead, he gives us various *existence proofs*, e.g. that if 0 is the number of an empty concept  $F$  and an empty concept  $G$ , then there must be a relation  $\phi$  bringing  $F$  and  $G$  under one-one correlation (1884, §75), or that there is something which immediately succeeds 0 (1884, §77). Equally, he justifies his diagnosis that the formalists have only introduced empty signs instead of new number-words on their failure to prove the existence of the new numbers (1884, §92ff.). The great importance of existence proofs comes out in the sharp contrast Frege draws, on several occasions, between defining a concept by means of the properties an object must have to fall under the concept and proving that something does fall under the concept (see 1884, §74, 1893, xiv.). This shows that definitions, on his account, can never bring objects into existence, but only specify concepts hitherto lacking a designator.

## 8 NUMBERS AS EXTENSIONS

Frege's project thus clearly has an ontological aim, to discover what is already there. Weiner's denial of this point even with respect to the *Foundations* is not convincing. She claims that there is no argument in the *Foundations* that numbers are really extensions, and points at §69 and §107 for evidence of this. Concerning §69, she seems to suggest that Frege avoids addressing the ontological question 'Are numbers extensions?', and asks instead a linguistic question, motivated by the context principle, namely whether the assertions which we make about extensions are assertions we can make about numbers. But Frege does not believe that in answering the linguistic question he is eschewing the ontological question, or dismissing it as out of place. On the contrary, his interest in language has an explicit ontological agenda: 'There is no intention of saying anything about the symbols; no one wants to know anything about them, except insofar as some property of theirs directly mirrors some property in what they symbolise' (1884, §24). The context principle, as formulated in the *Foundations*, is not employed as an anti-ontological tool, but is called to dispel the psychologistic prejudice that an expression can stand for an entity only when, taken in isolation, we associate a mental idea with it (1884, §59). Thus the context principle actually serves the ontological agenda: numbers are self-subsistent entities because their expressions can be construed as singular terms, not in isolation, but at least in the sentential contexts of their most paradigmatic arithmetical use, which is also ontologically significant (equations understood as identities).



Frege's remark that he does not attach decisive importance to bringing in extensions (1884, §107) is also not evidence against his ontologism. The opposite is true. Frege brings in extensions in the course of offering his 'explicit' definition of number. This definition is given expressly in response to the 'Julius Caesar' problem, namely that all recognition statements of the form ' $N_{\tau}F(\tau) = a$ ' must have a sense, i.e. that we must decide for any object  $a$  whether the statement is true or not (1884, §66-8, §107). But of course, recognition statements are identity statements, and their truth is a criterion for the objecthood of the content of the signs flanking the identity sign (1884, §62, §107). Hence, what Frege is most concerned with here is, again, an ontological issue: to specify, or at least sketch, a logicist criterion for the objecthood of numbers. He brings in extensions of concepts for this, which are logical objects on his account. The wariness he exhibits in §107 is therefore certainly not about the need to import *some* suitable objects to underpin his definition: such an import is essential, not indifferent to his definition of number. We can see this from the defence of his own suggestion that one could write 'concept' instead of 'extension of the concept' in his definition of number (1884, §68fn): by substituting 'concept' for 'extension of the concept' in the definition 'the Number which belongs to the concept  $F$  is the extension of the concept "equinumerous to the concept  $F$ "' the word 'concept' would be preceded by the *definite article*, and the whole phrase ('the concept') would be thus still a singular term, determining numbers as objects.<sup>26</sup> The wariness is rather about the fact that the phrase 'extension of the concept' is itself left undefined ('presupposed') in the *Foundations*, and hence that the 'Julius Caesar' problem remains open; for to decide whether  $N_{\tau}F(\tau) = a$  we will have to be able to decide sharply whether  $a$  is a certain extension. So Frege's wariness has an ontological motivation, that of securing a sharp objecthood criterion for numbers, which is not achieved just by employing the notion 'extension of the concept'. This interpretation is confirmed once we look at Frege's mature solution to the problem, as offered in *Grundgesetze*, where he brings in extensions as value-ranges, not as *defined*, but as *primitive* objects.<sup>27</sup> This is obviously an ontological move, moreover one entirely untouched by the stipulative role of definitions. Weiner's insistence on the stipulative role of definitions, as allegedly ruling out an ontological agenda, is out of focus.

<sup>26</sup> See Frege (1892, 48.) See also the illuminating discussion in (Burge 1984, 274-84.)

<sup>27</sup> See Dummett (1991, 159.)

## 9 CONCLUSION

I hope to have shown two things in this paper. First, Weiner's interpretation of the notions of 'Foundations-content', sense, reference and truth is extremely problematic. Second, the obvious requirement and the ontological thesis are very likely correct. Frege is a Platonic realist. He aims to define and analyse pre-existing arithmetical objects and concepts. A full defense of this view would have to look in more detail at the role of definitions in both *Foundations* and *Grundgesetze*, and the relation between the two, which cannot be done here.<sup>28</sup> In any case, I think it is more than probable, given the discussion above, that we have little chance to understand Frege's project of defining number, if we neglect his ontological agenda.

One final remark is called for. While Weiner is wrong in underplaying Frege's ontological agenda, she is surely on more firm ground in stressing the epistemological aim of his project. However, this also needs qualification. Weiner sees Frege realising the epistemological aim by means of a semantic undertaking, the sharpening of pre-systematic arithmetical language. But it is unclear how such a sharpening, by itself, would ever satisfy Frege's Cartesian craving for absolute certainty. Consider his repeated insistence that the *Foundations* have only established a probable thesis (1884, §87, §90), while his demand for total proof aims 'to place the truth of a proposition beyond all doubt' (1884, §2), to give it 'absolute certainty that it contains no mistake and no gap' (1884, §91), to raise the probability that arithmetical truths are analytic and *a priori* to a certainty (1884, §109) etc.<sup>29</sup> Clearly, vagueness of concepts is not the only source of doubt and error; a thinker might have sharpened all his concepts and still not be able to reach more than probable knowledge. 'X is a sharp concept, but it is uncertain whether y falls under X' is not incoherent. Adding the gapless proof requirement does not yield the desired certainty either, as we still need to access the unshakeable ground on which the derived propositions rest, the axioms expressing primitive truths (*Urwahrheiten, Urgesetze*, §2-4).<sup>30</sup> Frege has his eyes set on more than just increased conceptual and proof-technical rigour, to be achieved by mere stipulations. Instead, he formulates a programme of genuinely reductive analysis: an arithmetical truth has found its epistemological classification if we can trace its proof back to the primitive truths (1884, §4), whose number we have reduced to a minimum (1884, §2). Since primitive truths are truths evident without further proof, they must involve an

<sup>28</sup> See my forthcoming article investigating this.

<sup>29</sup> See also his talk about the 'unconditional assurance against a proof or a gap' (1884, §91fn.) and 'the secure ground under our feet' (1903, §62).

<sup>30</sup> Frege's simile is that of the 'Unerschütterlichkeit eines Felsblockes', which is best translated as 'unshakeability of a boulder'. This places Frege's metaphor in the gravitational orbit of Descartes' *fundamentum inconcussum*.

indubitable source of knowledge. Hence, Frege's epistemological project has a foundationalist and rationalist agenda. Moreover, there is no tension between Frege's foundationalism and his ontological agenda. On the contrary, the former presupposes the latter. The truths about logical objects are self-evident because of their nature: 'In arithmetic we are not concerned with objects which we come to know as something alien from without through the medium of the senses, but with objects given directly to our reason and, as its nearest kin, utterly transparent to it. And yet, or rather for that very reason, these objects are not subjective fantasies. There is nothing more objective than the laws of arithmetic' (1884, §105). In fact, Platonism is the basis of all knowledge: 'If there were nothing firm, eternal in the continual flux of all things, the world would cease to be knowable, and everything would be plunged in confusion' (1884, vii)<sup>31</sup>. At the ultimate level, epistemological questions are intimately bound with ontological ones.

#### REFERENCES

Note: *Begriffsschrift*, *Foundations* and *Grundgesetze* quotations always refer to the sections of the books. Frege's other published writings are cited by the original page. Posthumous writings are cited by the English translation in Gabriel et al. (eds.), 1979.

- Benacerraf, P., 1981, Frege: 'The last logicist'. *Midwest Studies in Philosophy* 6. 17-35.
- Burge, T., 1984, Frege on extensions of concepts. In Burge 2005.
- Burge, T., 1990, Frege on sense and linguistic meaning. In Burge 2005.
- Burge, T., 2005, *Truth, Thought, Reason: Essays on Frege*. Oxford, Oxford University Press.
- Carnap, R., 1910-1914/2004, *Frege's Lectures on Logic: Carnap's Student Notes*.
- Dummett, M., 1991, *Frege: Philosophy of Mathematics*. Cambridge/Mass., Harvard University Press.
- Frege, G., 1879, *Begriffsschrift*. Halle, Louis Nebert.
- Frege, G., 1879-1891, *Logik*. In Frege 1983.
- Frege, G., 1884, *Grundlagen der Arithmetik*. Breslau, Wilhelm Koebner.
- Frege, G., 1885, Über formale Theorien der Arithmetik. In Frege 1990.
- Frege, G., 1892, Über Begriff und Gegenstand. In Frege 1990.
- Frege, G., 1893, *Grundgesetze der Arithmetik I*. Jena. Verlag Hermann Pohle.
- Frege, G., 1899, Letter to Hilbert, 27.12.1899. In Gabriel et al. (eds.), 1980.
- Frege, G., 1903, *Grundgesetze der Arithmetik II*. Jena. Verlag Hermann Pohle.
- Frege, G., 1906a, Über die Grundlagen der Geometrie I. In Frege 1990.
- Frege, G., 1906b, Einleitung in die Logik. In Frege 1983.
- Frege, G., 1914, *Logik in der Mathematik*. In Frege 1983.
- Frege, G., 1918a, Der Gedanke. In Frege 1990.

<sup>31</sup> English translation amended by the author.

- Frege, G., 1918b, Die Verneinung. In Frege 1990.
- Frege, G., 1919, Aufzeichnungen für Ludwig Darmstaedter. In Frege 1983.
- Frege, G., 1924-1925, Sources of knowledge of mathematics and the mathematical natural sciences. In Frege 1983.
- Frege, G. 1983, *Nachgelassene Schriften*. Hamburg, Felix Meiner.
- Frege, G. 1984, *Collected Papers on Mathematics, Logic, and Philosophy*. Oxford, Basil Blackwell.
- Frege, G. 1990, *Kleine Schriften*. Hildesheim, Georg Olms Verlag.
- Gabriel, G., Hermes, H., Kambartel, F., Thiel, C., and Veraart, A. (eds.), 1980, *Philosophical and Mathematical Correspondence*. Chicago, University of Chicago Press.
- Geach, P. and M. Black (eds.), 1980, *Translations from the Writings of Gottlob Frege*. Oxford, Basil Blackwell.
- Hermes, H., F. Kambartel, F. Kaulbach (eds.), 1979, *Posthumous Writings*. Chicago, University of Chicago Press.
- Kemp, G. 1996, Frege's Sharpness Requirement. *Philosophical Quarterly* 46. 168-84.
- Picardi, E. 1988, Frege on definition and logical proof. In Cellucci C., Sambin, G. (eds.), *Temi e Prospettive della Logica e della Filosofia della Scienza Contemporanea*, vol. i. Bologna, Cooperativa Libreria Universitaria Editrice Bologna.
- Rumfitt, I. 2003, Singular terms and arithmetical logicism. *Philosophical Books* 44. 193-219.
- Shieh, S. 2008, Frege on definitions. *Philosophical Compass* 3/5. 992-1012.
- Stekeler-Weithofer, P. 1986, *Grundprobleme der Logik: Elemente einer Kritik der formalen Vernunft*. Berlin/New York. Gruyter.
- Weiner, J. 1984, The philosopher behind the last logicist. In Wright (ed.), 1984.
- Weiner, J. 1990, *Frege in Perspective*. Ithaca NY, Cornell University Press..
- Weiner, J. 2007, What's in a numeral? Frege's Answer. *Mind* 116. 677-716.
- Wright, C. (ed.), 1984, *Frege: Tradition and Influence*. Oxford, Basil Blackwell.

## The Indispensability of Logic

**Abstract.** The paper discusses the currently prominent strategy of justifying our elementary logical-inferential practices by their unavoidability and global indispensability for all our cognitive efforts. It starts by agreeing with prominent apriorists about their attempt to justify such beliefs based on constitutiveness (Boghossian) or based on the Global Indispensability Argument (C. Wright), and then proceeds to argue that unavoidable and indispensable tools provide entitlement/justification for projects if those projects are themselves meaningful. However, we are justified to think that our most general cognitive project is meaningful, and justified partly on the basis of its success up until now; and this basis is *a posteriori*. Therefore, the whole reflective justification from compellingness and unavoidability is *a posteriori*. This suggests that the justification of our intuition-based armchair beliefs and practices in general is plural and structured, with *a priori* and *a posteriori* elements combined in a complex way. It seems therefore that *a priori*/*a posteriori* distinction is useful and to the point. What is needed is refinement and respect for structure, not rejection of the distinction.

### 1 INTRODUCTION

How is simple and naïve logical reasoning justified? If, in the case of naïve cognizer, we distinguish the immediate entitlement, normally not consciously available to her, and in this sense 'external' to her, from reflective justification reserved for more sophisticated cognizers, our question branches into two. First, where does the entitlement of a naïve thinker who spontaneously and unthinkingly uses Modus Ponens or Conjunction Elimination come from? And when the naïve thinker becomes more sophisticated, where does her reflective justification come from? One powerful family of arguments for ultimate justification of our logical practice(s) concerns the following facts: simple rules of logic are compelling and unavoidable for humans, they enable the very having of

beliefs and constitute the rationality of reasoning. Because of this, they are both unavoidable and indispensable for our thinking, and for any sort of cognitive projects we might engage in. This secures the entitlement for the naïve reasoner, and justification for the sophisticated, reflective thinker. The apriorists add that these two, entitlement and justification, are a priori. This line of argument and the resulting family of arguments is extremely popular in contemporary debates, represented by leading thinkers on the matter, such as C. Wright, P. Boghossian, and to some extent P. Horwich and C. Peacocke (see References).

I shall assume here that the above line of argument is plausible indeed, but will argue that it is as yet incomplete and demands a further step that leads it away from the conclusion these thinkers prefer. I shall point out that *unavoidable and indispensable tools provide entitlement/justification for projects only if projects are themselves meaningful, and provide reflective justification only if we are also justified in finding them meaningful. However, we are justified to think that our most general cognitive project is meaningful partly on the basis of its success up until now; and this basis is available only a posteriori; we are justified in trusting that these projects are meaningful a posteriori, because they seem to have worked decently well up until the present. Therefore, the justification from compellingness and unavoidability is a posteriori, both in its role of warrant, usually spontaneously had by the naïve thinker, and in its role of reflective justification, found out and deployed by philosopher(s).*

The line of thought to be proposed in this paper has important similarities and dissimilarities with the traditional Quinean argument for indispensability. The similarity is in the basic appeal to a posteriori considerations of indispensability and success. The difference is structural: first, it is only one line among several, and it does not exclude some prima facie justification by obviousness and compellingness which is a priori; second, it concerns either the entitlement that is normally external and not present to the cognizer's awareness, or, if she is a reflective thinker, then also her high-level reflective justification.

In this paper I will present this line of argument in more detail, using as my foil the work of Boghossian and C. Wright. I apologize for not going into a longer debate with them, for lack of space. Let me start.

Our topic is the ordinary, naïve use of logic in reasoning. So, take a naïve reasoner N who passes from accepting a conjunctive statement to accepting one of its conjuncts (and acting upon her belief in it), thereby performing (what we would describe as) a step of elimination of conjunction. This can be brought to light by explicitly asking her a question about  $p$ -and- $q$  but not- $p$  situation, e.g. whether it is possible that the whole (conjunctive) statement holds, i.e. that the complex situation obtains, without things being as the relevant conjunct describes them. The "Of course, not" answer would confirm the impression that she does have a mastery of the rule governing conjunction. Call this knowledge instance-knowledge. It seems that knowledge manifested in such spontaneous inferences is knowledge how. However, when a naïve thinker begins to reflect,

the inferential step seems obvious, at least in its concrete implementation: the impossibility of the combined concrete situation is obvious, vividly presented to her. Thus she finds the situation in which it is the case that  $p$ -and- $q$  but not- $p$  (for some instances of " $p$ " and " $q$ ") inconceivable.

Let us follow common sense in assuming that N is making no mistake in reasoning the way she does. Let us refine this assumption by ascribing to N an immediate entitlement that she would find difficult to formulate; part of this entitlement stems from the fact that Conjunction elimination is in fact a valid rule, and part of it, we shall argue (along with our apriorists) comes from the indispensability of such steps for N's reasoning, cognitive project(s) and action. N could learn logic and epistemology and become reflectively aware of the issues involved; call N at that stage N\*. Or, an epistemologist might start reflecting upon N's (or her own) performance. They would then pass to a level of reflective internal justification.

In principle we thus have two directions of reflection: 1<sup>st</sup> person type, exemplified by our sophisticated N\* character, and 3<sup>rd</sup> person type, exemplified by our epistemologist. It would be interesting to explore similarities and differences between the two, but I must be short. Let me just state my view: 1<sup>st</sup> person reflective justifiedness is necessary for being completely justified. I am using as my framework a two-level picture of justification that I have picked up from the work of Russell and Sosa (see their works in References, and also the paper by Tom Baldwin listed there); but I hope that the results of the discussion can be easily applied, *mutatis mutandis*, to other frameworks; for instance, if you believe only in conscious, reflective justification, the result will be relevant for your framework as well. On the more externalist side, even if you disagree with my line on the importance of 1<sup>st</sup> person reflective knowledge, you might think that completely blind spontaneous thinking is not sufficient for justification. Only the most simple reliabilists think it is. Others think that some sort of availability of the 3<sup>rd</sup> person type justification is needed. (Peacocke and Boghossian might be good examples). For all of them, the issue of indispensability is crucial. So, what is exactly the role of indispensability, and what is its character; is the justification it bestows a priori or a posteriori?

## 2 THE INDISPENSABILITY ARGUMENT(S)

Here is then the preliminary sketch of my main argument:

- (0) Logical practice and beliefs stand in need of entitlement and justification.
- (1) Simple rules of logic are indispensable for humans, therefore

- (2) they are both unavoidable and indispensable for our thinking, and for any sort of cognitive projects we might engage in.
- (3) Unavoidable and indispensable tools provide entitlement/justification for projects iff the projects are themselves meaningful.
- (4) Our most general cognitive project has been at least minimally successful, and therefore it is meaningful and we are justified in believing that it is, and the naïve thinker is entitled to her logical reasoning.
- (5) This justification and entitlement are to a large extent a posteriori.

We now look briefly at each step. First, premise (0): the need for entitlement and justification. By entitlement for A-ing we mean that the A-er is permitted to A and is blameless in doing so. We also allow for first-level justification: obviousness and immediate compellingness does provide some good reason, but not all obvious and compelling beliefs and procedures are justified. Nash once said that he is receiving messages from extraterrestrials which come to him with the same degree of persuasiveness (and he probably meant obviousness and immediate compellingness) as his mathematical theorems. Since this phrase is printed on the cover of the paperback edition of the “Beautiful Mind”, let me call the problem The Beautiful Mind Problem. It introduces the need for second-level, reflective justification: why do I trust my use of Conjunction elimination, if geniuses like Nash found their logical reasoning as persuasive as messages from extraterrestrials? It is a more moderate problem than the related extreme Cartesian Problem of Madness which C. Wright uses as part of the motivation for his proposal for the appeal to indispensability.

In a discussion at Tim Williamson’s lecture in Dubrovnik (in the summer of 2009) he formulated an interesting point against The Beautiful Mind Problem: if you concede this much to the skeptic and you let the problem in, it will never get out! However, The Beautiful Mind Problem is not as global as the Madness problem, although the skeptic might try a slippery slope ascent from the former to the latter. Instead, the former merely explores and exploits our occasional doubts that are neither pathological nor, to my mind exaggerated. And we don’t have to convince the extreme skeptic, just come up with reasons we find good enough for trusting our reasoning faculties and strategies. So, we can uphold

- (0) Logical practice and beliefs stand in need of entitlement and justification.

We now pass to the next two premises, that contains the main point of the first part of our argument, stages (0)–(3), namely the one about indispensability. We shall discuss (1) and (2) together.



- (1) Simple rules of logic are indispensable for humans.
- (2) They are indispensable for our thinking, and for any sort of cognitive projects we might engage in.

So let me comment on (1) and (2) and summarize the extant arguments for them.

Let me start from what we might dub The Constitutiveness Argument, or Constitutiveness variant of the Indispensability Argument, as stated by Boghossian in his “Knowledge of logic” and developed in the recent book (2008). Then we shall pass to C. Wright’s Global Indispensability argument (2004) (the name for the argument is ours!). Boghossian’s proposal develops around the idea that logic can be justified in a rule-circular manner, due to its indispensability for thinking almost any contents whatsoever. Without dispositions to reason in accordance with logic, we could not even have the general belief whose justification is supposed to be in question, i.e. the belief about inferential potentials of a given logical constant. In a nutshell, the argument can be reconstructed as follows:

Certain of our inferential dispositions fix what we mean by our logical words (in the language of thought), therefore

without those dispositions there is nothing about whose justification we can even intelligibly raise a question.

Moreover, without those dispositions we could not even have the general belief whose justification is supposed to be in question. Therefore

We are entitled to act on those inferential dispositions prior to, and independently of, having supplied an explicit justification for the general claim that they are truth-preserving.

Here is the brief background story. The chapter “Epistemic Analyticity: A Defense”, of Boghossian’s (2008) book, starts with a piece of self-criticism: “at the time of writing that paper (i.e. “Analyticity Reconsidered”—NM), I did not delineate sufficiently clearly the difference between inferential and constitutive construals of the relation between meaning and entitlement” (Boghossian 2008, 225). The distinction is crucial for the author’s more recent work. Boghossian notes that the Implicit Definition Template involves a lot of inferring that already uses logic. So this premise-and-derivation model can neither entitle the reasoner to the crucial premise (3), i.e. that  $S(f)$  is true, nor explain what her entitlement to reason according to certain deductive rules consists in, since the entitlement presupposes such reasoning. We therefore need a contrasting model, and its crucial point is that the mere fact that the thinker grasps  $S$ ’s meaning entails that the thinker is justified in holding  $S$  to be true. (The epistemological consequences of the proposal are then developed in the next chapter, “How are objective epistemic reasons possible?”.) Take conditionals. If I don’t follow

Modus Ponendo Ponens (MPP), I can't have if-thoughts at all. So, if I do follow it, with  $p$  and 'if  $p$  then  $q$ ' as my premises, I cannot be blamed, so, I am entitled to follow it.

If inferring from those premises to that conclusion is required if I am to have the ingredient propositions, then, as a matter of metaphysical necessity, I cannot so much as consider the question whether the inference is justified without being disposed to reason in that way. Under those circumstances, then, it looks as though inferring according to MPP cannot be held against me, even if the inference is, as I shall put it, blind-unsupported by any positive warrant (Boghossian 2008, 230).

The chapter concludes by stressing that according to the "Constitutive model" the most fundamental relation between grasp of meaning and entitlement occurs when a thinker is entitled to reason in accordance with a certain rule simply by virtue of the fact that this rule is constitutive of a concept of his. The author expresses his hope that the model can be extended from reasoning to beliefs, if they are similarly constitutive of the possession of a concept (which has to be non-defective—we shall come to this in a moment). He proposes that this will solve the issue, famously raised by Aristotle (in *Metaphysics* Γ), about our entitlement to accept the principle of non-contradiction.

Boghossian has been developing the first line, on meaning-constitutiveness as the main *a priori* justifier, combined in his "Knowledge of Logic" with occasional remarks on compellingness, i.e. on the alleged fact that "it is not open to us to regard our fundamental logical beliefs as unjustifiable." (Boghossian and Peacocke 2000, 253) For instance, in the same paper Boghossian argues for the warrantedness of logical rules mainly based on negative compulsion, i.e. from deeply felt unacceptability and inconceivability. He does not offer any causal or psychological explanation of compulsion, which is, after all, a felt item. Here is the relevant quote:

[W]e cannot accept the claim that we have no warrant whatsoever for the core logical principles. We cannot conceive what such a warrant could consist in [...] if not in some sort of inference using those very core logical principles.

And further down:

It is not open to us to regard our fundamental logical beliefs as unjustifiable. (Boghossian 2000, 253.)

Another variant of the appeal to indispensability is the Global Indispensability Argument due to Crispin Wright. He revives the Wittgensteinian conception of hinges, generalizes it and enriches it with his own idea of "cornerstones":

Call a proposition a cornerstone for a given region of thought just in case it would follow from a lack of warrant for it that one could not rationally claim warrant for *any* belief in the region. The best—most challenging, most interesting—sceptical paradoxes work in two steps: by (i) making a case that a certain proposition (or restricted type of proposition) that we characteristically accept is indeed such a cornerstone for a much wider class of beliefs, and then (ii) arguing that we have no warrant for it. (Wright 2004, 167-8.)

If a cognitive project is “rationally non-optional”, i.e. indispensable in rational inquiry and in deliberation, then we may rationally take for granted the original presuppositions of such a project without specific evidence in their favor. The absence of defeating information is sufficient. So, elementary logic is both unavoidable and indispensable.

None of the arguments is final, as no philosophical argument is. But all of them converge on unavoidability and indispensability. If you believe in constitutive conceptual connections, then the Constitutivity Argument might appeal to you. And if you find Wittgenstein most congenial, on some of many readings of his text, then the Global Indispensability Argument will probably convince you. So, there are good reasons to accept the claim above.

### 3 ONLY A GOOD PROJECT JUSTIFIES ITS MEANS

We now pass to the second part of the argument, and move on to the particular twist we want to give to it, against its apriorist reading. Let us start with

- (3) Unavoidable and indispensable tools provide entitlement/justification for projects iff the projects are themselves meaningful.

The premise encapsulates the commonsense wisdom that no justification can come from bad, impossible and/ or idiotic projects! E.g., imagine a beginner who reasons: “If we want to square the circle, we need theorem  $\Theta$ ; therefore  $\Theta$ .” He would be very quickly taught that the project is impossible, so the theorem needed for it cannot be justified by the need. Next, consider clearly impossible lifelong projects: I want to achieve, by exercise, the height of 12 ft. so I am rational in doing the exercise. And finally, morally bad projects also demand lots of means: if one wants to build a torture chamber, one needs electricity at the least. So, one would desire some electricity for this purpose. Does indispensability of electricity morally justify the instrumental desire? Not really.

Corine Besson has objected (at a talk in Geneva, winter 2010) that even bad projects yield some instrumental justification for their means (thanks, Corine).

It seems to me that such a justification is conditional: if the project P is a good one, *then* means M is justified. It cannot be detached, and used independently, until P is independently justified.

Let us now consider the application of this reasoning to cognitive projects. Here is a story about *Mr. Magoo*.

Mr. Magoo has a very defective cognitive apparatus. His inductive propensities are idiotic, to use politically incorrect vocabulary, his senses most often deceive him, and his “heuristics” are ridiculous. (He lives in a super-hospitable environment, but hardly manages to survive.) His idiotic inductive propensities and ridiculous “heuristics”, plus his misplaced uncritical trust in his senses are indispensable for his ever forming any belief. Therefore, he is warranted in taking them as unquestioned and unquestionable starting points.

Of course, he is not warranted, most people would say, he is just being stupid. We can derive an argument from this kind of reaction: if you don't find Mr. Magoo's reasoning convincing this suggests that the justification of means for a cognitive project depends on the meaningfulness of the project itself. If the project has no chances to succeed, if *it* is hopelessly flawed (in the given context), then *it* is not justified. If it is not justified, it cannot lend its justifiedness to the means, since it does not have any.

This is valid both for entitlement and would be valid for reflective justification, if Mr. Magoo were capable of producing one. He is not entitled to his propensities and heuristics, nor to his trust in his senses. And if he were able, per impossibile, to produce a piece of reflective attempt at self justification, that would fail as well.

Let us apply this “Mr. Magoo Argument” to Crispin Wright's idea of cornerstones for a project. It seems that the acceptance of hinges and cornerstones is justified by the quality of cognitive projects they enable, and is sensitive to the chances of their success. But these chances are revealed by trying. Therefore, our best access to our own warrant involves information about the success of the relevant cognitive project. The warrant for logic is thus sensitive at least to the chances of success of our “total inquiry”, and our awareness of it depends on the information about the success. Meaningfulness is not independent of chances of success, in virtue of “ought implies can” principle.

Would this make the final justification merely pragmatic? This is a question that I have often heard at talks. When meant as a criticism, I think it rests upon confusion. The justification is instrumental, but the goal appealed to is truth, reliability or some such epistemic value, so being instrumental does not make it into a pragmatic justification in a non-epistemic, merely practical sense. “Success” here is epistemic success, not practical-pragmatic. Alternatively, if

“pragmatic” just means instrumental, then the question does not entail criticism: there is nothing wrong with a piece of belief (or cognitive habit) being justified by its contribution to the achievement of epistemic value(s). So we come to the two final steps:

- (4) Our most general cognitive project has been at least minimally successful; therefore, it is meaningful and we are justified in believing that it is, and the naïve thinker is entitled to her logical reasoning.
- (5) This justification and entitlement are to a large extent a posteriori.

Now, the success of our “total inquiry” is to a large extent an empirical matter. Therefore, our awareness of it depends, to a great extent, on empirical information. Such information is a posteriori. Consider reflective justification: how does the cognizer arrive at justified beliefs about herself being warranted? Well, partly by relying on relative success of her total project. And this reflective justification might therefore be seriously a posteriori, in a way that precludes purely a priori justification of logic.

But you need deductive logic from the start of the inquiry, and you need inductive-logical assumptions for evaluating your empirical evidence; so you can't get rid of the a priori, the objector might argue. Not really: some assumptions may be pragmatically antecedent to a cognitive project, e.g. those that will later be classified as logical, but they are, firstly, reflectively justified by the overall success-chances of the project, and secondly, revisable in the light of some advanced stage of the project. We can thus conclude: *both first-order entitlement and reflective justification from indispensability are a posteriori.*

But aren't we back to the old Quine-Putnam indispensability? A comparison is needed. First, the old indispensability has been presented as replacing obviousness and compellingness, and second, it was implicitly presented as first-order property of our beliefs. This has made it vulnerable to quite dangerous objections: that it bypasses well-entrenched and extremely stable traditional justifiers, that it does not correspond to the first-order intuitions of practitioners themselves (for instance, mathematicians who just accept obvious-looking moves and their results). Our proposal avoids these pitfalls: it is compatible with a prima facie role for obviousness, and it situates the appeal to success by the thinker herself into the lofty spheres of reflective self-understanding where it belongs.

## 4 CONCLUSION: FROM INDISPENSABILITY TO APOSTERIORITY

Much more needs to be said, but let us summarize how far we have come: unavoidable and indispensable tools provide entitlement/justification for projects if projects are themselves meaningful.<sup>1</sup> However, we are justified to think that our most general cognitive project is meaningful, and justified partly on the basis of its success to date; and this basis is a posteriori. There is more than just a touch of aposteriority present in the considerations of meaningfulness, and much more in our coming to know about our warrant: reflective justification based on compellingness and unavoidability is wholly a posteriori. This suggests that the justification of our intuitional armchair beliefs and practices in general is plural and structured, with a priori and a posteriori elements combined in a complex way. It seems therefore that the a priori/ a posteriori distinction is useful and to the point.

Some philosophers suggest that we instead drop the a priori/ a posteriori contrast. (A. Goldman claims that warrant is just a complex and multi-dimensional affair (1999, 48), so the contrast is misplaced. T. Williamson argues for the same conclusion in his recent (2007) book. It is a bad idea: we need to distinguish and recognize structure, rather than obscure it. For instance, the immediate justification of logical moves is certainly a priori. But the reflective justification is not. What is needed is refinement and respect for structure, not rejection of the distinction. But, if there is structure, what is the final verdict? Is justification of simple logical moves and beliefs a priori or a posteriori? A traditional principle insisted on purity of the a priori: if justification (or entitlement) contains a posteriori elements, then it is ultimately a posteriori (e.g. if it is mixed and contains one a posteriori element, it is ultimately a posteriori). So, if you accept the principle, you might talk about structured a posteriority. If not, you would merely talk about structured justification and entitlement.

To conclude and reiterate: many prominent apriorists have given up on focusing their defense of knowledge of logic on traditional internalist grounds of obviousness, self-evident character and the like. Instead, they have revived the Indispensability Argument in a sophisticated setting, appealing to entitlement and reflexive justification, hoping to cleanse the argument from its Quinean a

<sup>1</sup> What about primitive compulsion as an alternative? It would solve the justification problem by “ought implies can” and offers some guidance, but opens the problem of being primitively compelled to hold true sentences (and to hold correct rules) one only partially understands. Worse, it offers a competing explanation: it is not stipulation per se but its irresistible force. (In addition, it is hard to see how a set of sentences does not yield analytic propositions by merely being held true, but does yield by being irresistibly held true .) There is a possibility: retreat to anthropocentrism: well, these are *our* concepts. Irresistibility is the mark of conceptual character, and irresistibility-for-us is the mark of ownership, of being “our own” concepts. But this move is not opened to Boghossian, since it would open the door to epistemic relativism, a view he is justly combating.

posteriorist heritage. I have argued that in this very setting the Indispensability Argument brings a posteriority back, since the issue of both entitlement and of reflective justification of logical and elementary mathematical beliefs and inferential propensities is to be decided to a large extent on the basis of the global successfulness of our cognitive effort, which is largely an a posteriori matter. To a significant degree, logic and elementary mathematical understanding are reflectively justified in an a posteriori manner. This should prompt us to opt for a more structured view of justification, containing essentially a posteriori elements, but in a different form than in the Quinean tradition.\*

## REFERENCES

- Baldwin, T., 2003, From knowledge by acquaintance to knowledge by causation. In Griffin, N. (ed.), *Cambridge Companion to Bertrand Russell*. Cambridge, Cambridge University Press.
- Boghossian, P., 1996, Analyticity. In Hale, B. and Wright, C. (eds.), *A Companion to the Philosophy of Language*. Oxford, Blackwell.
- Boghossian, P., and Peacocke C. (eds.), 2000, *New Essays on the A Priori*. Oxford, Clarendon Press.
- Boghossian, P., 2000, Knowledge of logic. In Boghossian and Peacocke (2000).
- Boghossian, P., 2001, How are objective epistemic reasons possible?. *Philosophical Studies* 106. 1–40. Reprinted in Boghossian, 2008.
- Boghossian, P., 2003, Blind reasoning. *Proceeding of the Aristotelian Society* Supplementary Volume 77. 225-48.
- Boghossian, P., 2008, *Content and Justification*. Oxford, Oxford University Press.
- Goldman, A., 1999, A priori warrant and naturalistic epistemology. *Philosophical Perspectives* 15.
- Horwich, P., 2000, Stipulation, meaning, and apriority. In Boghossian and Peacocke (2000).
- Boghossian, P. and Peacocke C. (Eds.), 2000, *New Essays on the A priori*. Oxford, Oxford University Press.
- Peacocke, Ch., 2004, *The Realm Of Reason*. Oxford, Oxford University Press.
- Sosa, E., 1991, *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge, Cambridge University Press.
- , 2007, *A Virtue Epistemology—Apt Belief and Reflective Knowledge*. Volume One. Oxford, Oxford University Press.
- Russell, B., 1926, *Our Knowledge of the External World as a Field for Scientific Method in Philosophy*.<sup>2</sup> Chicago and London, Allen and Unwin.
- , 1927, *An Outline of Philosophy*. London, Allen and Unwin.
- , 1940/1968, *An Inquiry into Meaning and Truth*. London, Allen and Unwin (paperback edition).
- , 1948, *Human Knowledge: Its Scope and Limits*. London, Allen and Unwin.
- Williamson, T., 2007, *Philosophy of Philosophy*. Oxford, Blackwell.
- Wright, C., 2004, On Epistemic Entitlement: I. Warrant for Nothing (and Foundations for Free)? *Proceeding of the Aristotelian Society*, 167-212.

\* I wish to thank the organizers of ELTE conference, and the participants in the discussion.

# Fitch's paradox and Labeled Natural Deduction System

**Abstract.** This paper introduces a relatively novel system of representing modal logic in a form of natural deduction. It then expands it to accommodate the epistemic operator and applies it to generate a more precise formulation of Fitch's paradox of knowability. Finally, an illustration of the paradox's pertinence to contemporary philosophical debate is laid out.

## 1 INTRODUCTION

The purpose of this paper is to provide the means for presenting Fitch's paradox, a philosophical argument requiring multiple modalities, within a purely formal deduction system. A labeled natural deduction system for modal logic offered by David Basin, Sean Matthews and Luca Vigano, provides the basis which is then expanded to accommodate an epistemic operator "know." An advantage of this system: anyone familiar with first-order natural deduction is provided with the means to formulate a useful and fruitful philosophical argument in a more precise manner at no added complexity.

The remainder of this section will lay out some desirable properties of any natural deduction system that we will naturally strive to meet in this paper. The second section introduces the labeled natural deduction system of Basin *et al.* and expands on it to allow us to formulate Fitch's paradox. Note that, while the authors use the "Gentzen-style" form of representing natural deduction, due to the relative length of the argument and the number of assumptions needed, for ease of presentation, the form used here is the "Suppes-Lemmon style." The third section presents Fitch's paradox first in an informal, and then in a formal manner. Finally, the fourth section provides the summary.



### 1.1 *Natural deduction systems*

Although natural deduction was first developed 1934, it is partially based on a proposal put forth in 1926 by Jan Łukasiewicz, who called for a system that can yield the same theorems as the axiomatic systems of the time, but which would follow more closely the actual practice of constructing a proof<sup>1</sup>. This system, reflecting the “natural” way humans reason, would follow where “arbitrary assumptions” lead and how long they stay in effect. This desirable property is something to keep in mind while presenting a natural deduction system.

In their widely used logic handbook *Language, Proof and Logic* Jon Barwise and John Etchemendy state that: “... [natural deduction] systems are intended to be models of the valid principles of reasoning used in informal proofs.”<sup>2</sup> This is precisely the purpose of the natural deduction system they present.

## 2 LABELED NATURAL DEDUCTION SYSTEM FOR MODAL LOGIC

This section introduces a labeled natural deduction system developed by Basin *et al.* Following the ideas of Dov Gabbay, this system provides a framework for capturing a large number of non-classical logics<sup>3</sup>. The focus here will be on modal logic. The peculiarity of the system is that it introduces a set of labels  $W$ ,  $W = \{x_0, x_1, \dots, x_n, \dots, y_0, y_1, \dots, y_n, \dots\}$ . These labels can be thought of as representing worlds in a Kripke model. The language of the labeled natural deduction system (henceforth: LNDS) differs from the standard, and widely familiar, (propositional) modal logic language precisely with regard to  $W$ .

### 2.1 *The language of LNDS*

The language of LNDS comprises two types of formulas, labeled and relational well-formed formulas. The latter concern the relations of labels, and correspond to the properties of the accessibility relation  $R$  in a Kripke model, whereas the former are merely modal propositional well formed formulas expanded with a label; we will define these first, and proceed from there. The (inductive) definition of a modal  $\mathfrak{wff}$  should be familiar:

<sup>1</sup> (Pelletier, F., 2000).

<sup>2</sup> (Barwise, J., Etchemendy, J., 2003).

<sup>3</sup> (Basin & *al.* 1998).

*Definition 2.1: Modal wff*

- 1 Propositional letters  $P$ ,  $Q$  and  $R$  are well formed formulae (*wff*).
- 2 If  $P$  is a *wff*, then  $\neg P$  is a *wff*.
- 3 If  $P$  is a *wff* and  $Q$  is a *wff*, then  $(P \wedge Q)$  is a *wff*.
- 4 If  $P$  is a *wff* and  $Q$  is a *wff*, then  $(P \vee Q)$  is a *wff*.
- 5 If  $P$  is a *wff* and  $Q$  is a *wff*, then  $(P \rightarrow Q)$  is a *wff*.
- 6 If  $P$  is a *wff* and  $Q$  is a *wff*, then  $(P \leftrightarrow Q)$  is a *wff*.
- 7 If  $P$  is a *wff*, then  $\Box P$  is a *wff*.
- 8 If  $P$  is a *wff*, then  $\Diamond P$  is a *wff*.
- 9 Nothing else is a *wff*.

Now we have all the ingredients necessary to define a *labeled well-formed formula*:

*Definition 2.2: Labeled well-formed formula (lwff)*

Let  $P$  be a modal *wff* (Def. 2.1), let  $W$  be a set of labels  $W = \{x_0, x_1, \dots, x_n, \dots, y_0, y_1, \dots, y_n, \dots\}$ , and let  $x$  be a member of such a set,  $x \in W$ . Then  $x:P$  is a *lwff* which can be understood as meaning “ $P$  is the case in (a possible world)  $x$ .”

As noted earlier, a *relational wff* is concerned with a relation of two labels:

*Definition 2.3: Relational well formed formula (rwff)*

Let  $x$  and  $y$  be members of a set of labels  $W$  (as above). Then  $xRy$  is a *rwff* which can be understood as “ $y$  is accessible to  $x$ .”

## 2.2 Rules of inference

In this section we will explore the rules of inference in LNDS. The rules of inference for truth-functional connectives should be readily recognizable to anyone familiar with natural deduction—the only novelty here being that each line is expanded with (one and the same) label. The rule for negation introduction deviates from this pattern, and will be discussed separately. Afterwards, rules of inference for modal operators will be covered, along with examples to illustrate them. The mode of presentation is “Suppes–Lemmon” style—the central column contains the enumerated steps of the proof, assumptions or derived formulas. The column on the right contains the “justification” of a step—a rule of inference used, or “ $P$ ” if it is an assumption (“ $P^*$ ” denotes an additional assumption that needs to be discharged, and the column on the left contains a set of undischarged assumptions the step “relies” on (assumptions “rely” on themselves). In a general form laid out here, the letters  $m, n, i, j, \dots$  signify numbers, Greek letters  $\Gamma$  and  $\Delta$  signify sets of assumptions, and letters  $A, B, \dots$  signify *wffs*.

## 2.2.1 Truth-functional connectives

Below are the rules for conditional and conjunction, which behave in a familiar way, the only difference being that each *wff* is expanded into a *labeled wff*, using the same label in each instance.

$$\begin{array}{l} \rightarrow \textit{Intro} \\ \{m\} \quad m \quad x:A \quad P^* \end{array}$$

$$\begin{array}{l} \Gamma \cup \{m\} \quad i \\ \Gamma \quad j \end{array} \quad \begin{array}{l} \dots \\ x:B \\ x:A \rightarrow B \end{array} \quad \rightarrow I: m, j$$

$$\begin{array}{l} \rightarrow \textit{Elim} \\ \Gamma \quad m \\ \Delta \quad i \\ \Gamma \cup \Delta \quad j \end{array} \quad \begin{array}{l} x:A \rightarrow B \\ x:A \\ x:B \end{array} \quad \rightarrow E: m, j$$

$$\begin{array}{l} \wedge \textit{Intro} \\ \Gamma \quad m \\ \Delta \quad i \\ \Gamma \cup \Delta \quad j \end{array} \quad \begin{array}{l} x:A \\ x:B \\ x:A \wedge B \end{array} \quad \wedge I: m, j$$

or

$$\begin{array}{l} \Gamma \cup \Delta \quad j \end{array} \quad \begin{array}{l} x:A \wedge B \end{array} \quad \wedge I: m, j$$

$$\begin{array}{l} \wedge \textit{Elim} \\ \Gamma \quad m \\ \Gamma \quad i \\ \text{or} \\ \Gamma \quad i \end{array} \quad \begin{array}{l} x:A \wedge B \\ x:A \\ x:B \end{array} \quad \wedge E: m$$

$$\begin{array}{l} \neg \textit{Intro} \\ \{m\} \quad m \quad x:A \quad P^* \\ \dots \\ \Gamma \cup \{m\} \quad i \\ \Gamma \quad j \end{array} \quad \begin{array}{l} y:\perp \\ x:\neg A \end{array} \quad \neg I: m, j$$

(Or, alternatively)

$$\begin{array}{l} \neg \textit{Intro} \\ \{m\} \quad m \quad x:\neg A \quad P^* \\ \dots \\ \Gamma \cup \{m\} \quad i \\ \Gamma \quad j \end{array} \quad \begin{array}{l} y:\perp \\ x:A \end{array} \quad \neg I: m, j$$

Note that not all the lines here contain the same label (the label used in the line ( $i$ ) is “ $y$ ”). An impossible result in one world (i.e. under one label) can “transfer” to another world—a fact that Basin *et al.* call a “global falsum.” This is elaborated in Section 2.4.

$\perp$ <i>Intro</i>			
$\Gamma$	$m$	$x:A$	
$\Delta$	$i$	$x:\neg A$	
$\Gamma \cup \Delta$	$j$	$y:\perp$	$\perp I: m, i$
$\perp$ <i>Elim</i>			
$\Gamma$	$n$	$x:\perp$	
$\Gamma$	$i$	$x:A$	$\perp E: n$

### 2.2.2 Modal operators

The novelty of this approach consists in the introduction of natural deduction rules for the modal operators “ $\Box$ ” (“necessarily”) and “ $\Diamond$ ” (“possibly”). Note that these rules of inference are analogous to the rules for universal and existential quantifiers, respectively (with an “arbitrary label” replacing an “arbitrary name”).

$\Box$ <i>Intro</i>			
$\{m\}$	$m$	$xRy$	$P^*$
$\Gamma$	$i$	$y:A$	
$\Gamma - \{m\}$	$j$	$x:\Box A$	$\Box I: m, i$

Note:  $y$  is a new label, such that  $x \neq y$ , and not appearing in any of the suppositions in  $\Gamma$ , except perhaps  $\{m\}$ .

$\Box$ <i>Elim</i>			
$\Gamma$	$m$	$x:\Box A$	
$\Delta$	$i$	$xRy$	
$\Gamma \cup \Delta$	$j$	$y:A$	$\Box E: m, i$

*Example 2.1:* See the Appendix.

$\diamond$ <i>Intro</i>			
$\Gamma$	$m$	$x:A$	
$\Delta$	$i$	$xRy$	
$\Gamma \cup \Delta$	$j$	$x: \diamond A$	$\diamond I: m, n$
$\diamond$ <i>E.lim</i>			
$\Gamma$	$m$	$x:A$	$P^*$
$\{i\}$	$i$	$xRy$	$P^*$
$\{j\}$	$j$	$x: \diamond A$	$\diamond I: m, n$
		$\dots$	
$\Delta \cup \{i\} \cup \{j\}$	$k$	$z:B$	
$\Gamma \cup \Delta$	$l$	$z:B$	$\diamond E: m, i, j, k$

Note:  $y$  is a new label, such that  $y \neq x$  and  $y \neq z$ , which does not appear in any of the suppositions from  $\Gamma$  and  $\Delta$ .

*Example 2.2:* See the Appendix.

### 2.3 Familiar axioms

As noted, it is expected of a natural deduction system that it provide the same results as an axiomatic theory. Therefore, what follows are proofs of two well-known modal axioms—the rule of necessitation, which states that every theorem is necessary, and axiom K, which demonstrates how the necessity operator “ $\square$ ” is distributed over conditionals.

*Proof 2.1:* Rule of Necessitation (R.N)

Let  $x:A$  be a theorem, and  $x$  an arbitrary label. Proof for  $x: \square A$  will proceed as follows:

$\{m\}$	$m$	$xRy$	$P^*$
		<i>the proof of a theorem where each occurrence of the label <math>x</math> is substituted for the label <math>y</math></i>	
$\{\}$	$n$	$y:A$	<i>from the preceding proof</i>
$\{\}$	$j$	$x: \square A$	$\square I: m, n$

*Proof 2.2: Axiom K*

{1}	1	$x: \Box (A \rightarrow B)$	$P^*$
{2}	2	$x: \Box A$	$P^*$
{3}	3	$xRy$	$P^*$
{1,3}	4	$y: A \rightarrow B$	$\Box E: 1,3$
{2,3}	5	$y:A$	$\Box E: 2,3$
{1,2,3}	6	$y:B$	$\rightarrow E: 4,5$
{1,2}	7	$x: \Box B$	$\Box I: 3,6$
{1}	8	$x: \Box A \rightarrow \Box B$	$\rightarrow I: 2,7$
{}	9	$x: \Box (A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$	$\rightarrow I: 1,8$

*2.4 Global falsum*

One consequence of the negation introduction rule is the rule called “global falsum”:  $\Gamma \vdash_{x:\perp} \Rightarrow \Gamma \vdash_{x:\perp}$ .<sup>4</sup>

*Proof 2.3: Global falsum (gf)*

Suppose that (1)  $\Gamma \vdash_{x:\perp}$ . Then  $\Gamma, y:P \vdash_{x:\perp}$  (adding a premise does not alter the validity of a valid argument). It follows by  $\neg I$  that (2)  $\Gamma \vdash_{y:\neg P}$ . But in the same way, from (1) we can derive  $\Gamma, y:\neg P \vdash_{x:\perp}$ , and another application of  $\neg I$  gives (3)  $\Gamma \vdash_{y:P}$ . Applying  $\perp I$  to (2) and (3) yields  $\Gamma \vdash_{y:\perp}$ .

Since  $x$  and  $y$  represent arbitrary labels, it is obvious that falsum can “travel” freely between labels. The reason for the inclusion of the rule *global falsum* is that it allows a desirable result—interchangeability of  $\Box$  and  $\neg \Diamond \neg$ .

*Global falsum*

$\Gamma$	1	$x:\perp$	<i>gf:m</i>
$\Gamma$	2	$x:\perp$	

The following two proofs demonstrate how this inference rule allows for the derivation of that desirable result.

<sup>4</sup> (Basin & al. 1998).

*Proof 2.4a*

{1}	1	$x: \Box A$	$P$
{2}	2	$x: \Diamond \neg A$	$P^*$
{3}	3	$y: \neg A$	$P^*$
{4}	4	$xRy$	$P^*$
{1,4}	5	$y:A$	$\Box E: 1,4$
{1,3,4}	6	$y:\perp$	$\perp I: 3,5$
{1,3,4}	7	$x:\perp$	$gf: 5$
{1,2}	8	$x:\perp$	$\Diamond E: 2,3,4,6$
{1}	9	$x: \neg \Diamond \neg A$	$\neg I: 2,7$

*Proof 2.4b*

{1}	1	$x: \neg \Diamond \neg A$	$P$
{2}	2	$xRy$	$P^*$
{3}	3	$y: \neg A$	$P^*$
{2,3}	4	$x: \Diamond \neg A$	$\Diamond I: 2,3$
{1,2,3}	5	$x:\perp$	$\perp I: 1,4$
{1,2}	6	$y:A$	$\neg I: 3,5$
{1}	7	$x: \Box A$	$\Box I: 2,6$

2.5 *Relational rules*

Relational rules have the general form  $t_1Rs_1 \dots t_mRs_m \vdash t_0Rs_0$ , where  $t_0, t_1, \dots, t_m, s_0, s_1, \dots, s_m$  are members of the set of labels  $W$ . Relational rules mirror properties of the accessibility relation, and allow us to derive the corresponding axioms. The only rule necessary for the construction of Fitch's paradox is the relational rule of reflexivity, and it is therefore the only one presented here.

*Reflexivity*

	$M$		
{}	$I$	$xRx$	<i>RefI:</i>

*Proof 2.5:* axiom T

{1}	1	$xRx$	<i>Rrefl:</i>
{2}	2	$x: \Box A$	<i>P*</i>
{2}	3	$x:A$	$\Box E: 1,2$
{}	4	$x: \Box A \rightarrow A$	$\rightarrow I: 2,3$

It is clear how this is in keeping with the historical requirement posed for natural deduction—to yield the same results as an axiomatic theory.

### 3 FITCH'S PARADOX

Fitch's paradox, also known as the paradox of knowability, first appeared in Fitch's 1963 article "*A Logical Analysis of Some Value Concepts.*" There, the paradox appears in Theorem 5, which states:

If there is some true proposition which nobody knows (or has known or will know) to be true, then there is a true proposition which nobody can know to be true.<sup>5</sup>

However, an equivalent claim, which states that if all truths are knowable, then all truths are known, is usually considered when discussing the paradox:

$$\forall p (p \rightarrow \Diamond Kp) \vdash \forall p (p \rightarrow Kp)$$

#### 3.1 Informal proof of the paradox

The strength of the paradox derives from the fact that it rests on mostly unproblematic principles. They are:

$$(KIT): Kp \vdash p,$$

which states that knowledge is factive, i.e. knowledge implies truth.

$$(K Dist): K(p \wedge q) \vdash Kp \wedge Kq,$$

which states that knowledge is distributed over conjunction, i.e. knowledge of a conjunction implies knowledge of the conjuncts.

<sup>5</sup> (Fitch, F., 1963).



Furthermore, we must rely upon the rule of necessitation (*RV*)—all theorems are necessary. The fourth and final principle states that if  $p$  is necessarily false, it is impossible:

$$(P4): \Box \neg p \vdash \neg \Diamond p$$

*Proof 3.1: Fitch's paradox*<sup>6</sup>

Now, suppose that every truth is knowable:  $\forall p (p \rightarrow \Diamond K p)$ . Suppose also that we are not omniscient, that there is a truth which is not known:  $\exists p (p \wedge \neg K p)$ . Let  $p$  be such a truth:

$$(1) p \wedge \neg K p$$

Now, since every truth is knowable, so is (1):

$$(2) (p \wedge \neg K p) \rightarrow \Diamond K (p \wedge \neg K p)$$

Therefore, by *modus ponens*,

$$(3) \Diamond K (p \wedge \neg K p)$$

This, however, can be proven to be false. Let us suppose (for *reductio ad absurdum*):

$$(4) K (p \wedge \neg K p)$$

It follows by *K Dist* that both conjuncts are known:

$$(5) K p \wedge K \neg K p$$

And, applying *KIT* to the second conjunct, we get a contradiction:

$$(6) K p \wedge K p$$

That allows us to negate (4):

$$(7) \neg K (p \wedge \neg K p)$$

And, since (7) is a theorem, we can apply *RV* to get:

$$(8) \Box \neg K (p \wedge \neg K p)$$

Applying the fourth principle, *P4*, we get the opposite of (3):

$$(9) \neg \Diamond K (p \wedge \neg K p)$$

<sup>6</sup> (Brogaard, B., Salerno, S., 2008).

Obviously this means there is no unknown truth

$$(10) \neg \exists p (p \wedge \neg K p)$$

Or, in other words, that all truths are known:

$$(11) \forall p (p \rightarrow K p)$$

So, supposing that all truths are knowable leads us, very convincingly, to the conclusion that all truths are, in fact, known.

### 3.2 Formal proof of the paradox

Obviously, for the proof of the paradox to be constructed, we need to have rules for the operator  $K$ . These rules will mirror the inference rules for the necessity operator (using a separate accessibility relation,  $R_E$ ) and will allow us to derive all the principles needed in the informal proof.

$K\ Intro$	$\{m\}$	$m$	$xR_E y$	$P^*$
			$\dots$	
	$\Gamma \cup \{m\}$	$i$	$y:p$	
	$\Gamma$	$j$	$x: K p$	$K I: m, i$
		$\cdot$		
$K\ Elim$				
	$\Gamma$	$m$	$x: K p$	
	$\Delta$	$i$	$xR_E y$	
	$\Gamma \cup \Delta$	$j$	$y:p$	$K E: m, i$

Additionally, the relational rule of reflexivity for  $R_E$  will be introduced. It insures the  $KIT$  principle in keeping with the proof of axiom  $\Gamma$  in *Proof 2.5*.

*Proof 3.2: K Dist*

{1}	1	$x: K(p \wedge q)$	$P$
{2}	2	$xR_E y$	$P^*$
{1,2}	3	$y: p \wedge q$	$KE: 1,2$
{1,2}	4	$y: p$	$\wedge E: 3$
{1}	5	$x: K p$	$KI: 2,4$
{6}	6	$xR_E z$	$P^*$
{1,6}	7	$z: p \wedge q$	$KE: 1,6$
{1,6}	8	$z: q$	$\wedge E: 7$
{1}	9	$x: K q$	$KI: 6,8$
{1}	10	$x: K p \wedge K q$	$\wedge I: 5,9$

*Proof 3.3: KIT*

{1}	1	$x: K p$	$P$
{}	2	$xR_E x$	$R_E \text{ refl:}$
{1}	3	$x: p$	$KE: 1,2$

Obviously, the remaining principles have already been proven— $R_V$  in the Proof 2.1, and  $P4$  in the Proof 2.4b, substituting  $\neg p$  for  $A$ .

*Proof 3.4: Fitch's paradox in LNDS*

The formal version of the proof starts out in the same way—assuming that  $p$  is an unknown truth (2), but that every truth is knowable, and thus  $p \wedge \neg K p$  as well as (1). Again, this leads to the claim that it is possible to know that something is an unknown truth (3). We now set out to prove (in line 23) that this not the case. Note that for the sake of legibility, the words *label* and *world* are used interchangeably.

{1}	1	$x: (p \wedge \neg K p) \rightarrow \diamond K(p \wedge \neg K p)$	$P$
{2}	2	$x: p \wedge \neg K p$	$P^*$
{1,2}	3	$x: \diamond K(p \wedge \neg K p)$	$\rightarrow E: 1,2$

We assume that there is an accessible world  $y$  in which it is known that  $p$  is an unknown truth:

{4}	4	$xR_A y$	$P^*$
{5}	5	$y: K(p \wedge \neg K p)$	$P^*$

However, in that case  $p \wedge \neg K p$  holds in  $y$ , and therefore,  $\neg K p$  holds in  $y$ . These steps correspond to an application of principles *K Dist* and *K IT*:

{}	6	$yR_E y$	$R_E refl$ :
{5}	7	$y: p \wedge \neg K p$	$KE: 5,6$
{5}	8	$y: \neg K p$	$\wedge E: 7$

At the same time, if  $p \wedge \neg K p$  is known in  $y$ , then  $K p$  holds in  $y$ .

{9}	9	$yR_E z$	$P^*$
{5,9}	10	$z: p \wedge \neg K p$	$KE: 5,9$
{5,9}	11	$z: p$	$\wedge E: 10$
{5}	12	$z: K p$	$KI: 9,11$

Lines 8 and 12 are contradictory—they imply that something is an unknown truth can not be known.

{5}	13	$y: \perp$	$\perp I: 8,12$
{}	14	$y: \neg K (p \wedge \neg K p)$	$\neg I: 5,13$

Since  $y$  is an arbitrary world, it is necessarily unknowable that something is an unknown truth. This step corresponds to the line (8) of the informal proof.

{}	15	$x: \Box \neg K (p \wedge \neg K p)$	$\Box I: 4,14$
----	----	--------------------------------------	----------------

Now we need to perform the transformation from line (9) of the informal proof. To do so, we will assume that the opposite holds:

{16}	16	$x: \Diamond K (p \wedge \neg K p)$	$P^*$
------	----	-------------------------------------	-------

Of course, if  $K (p \wedge \neg K p)$  is possible, then there is a world in which it is true:

{17}	17	$y: K (p \wedge \neg K p)$	$P^*$
{18}	18	$xR_A y$	$P^*$

But, even in that world,  $\neg K (p \wedge \neg K p)$  is true (since it is, according to line 15, necessary). This leads to a contradiction:

{18}	19	$y: \neg K (p \wedge \neg K p)$	$\Box E: 15,18$
{17,18}	20	$y: \perp$	$\perp I: 17,19$

That contradiction transfers back to the original world:

{17,18}      21                                       $x:\perp$                                       *gf*: 20

And so the assumption in line 16 proves, be false, as we had hoped.

{16}            22                                       $x:\perp$                                        $\diamond E$ : 16,17,18,21  
 {}              23                                       $x:\neg \diamond K(p \wedge \neg K p)$                                        $\neg I$ : 16,22

Finally, we have shown that there are no unknown truths:

{1,2}           24                                       $x:\perp$                                        $\perp I$ : 3,23  
 {1}             25                                       $x:\neg(p \wedge \neg K p)$                                        $\neg I$ : 2,24

Of course, since  $p$  is an arbitrary proposition, it can be shown that all truths are, in fact, known. This transformation is trivial:

{26}            26                                       $x:p$                                        $P^*$   
 {27}            27                                       $x:\neg K p$                                        $P^*$   
 {26,27}       28                                       $x:p \wedge \neg K p$                                        $\wedge I$ : 26,27  
 {1,16,27}     29                                       $x:\perp$                                        $\perp I$ : 25,28  
 {1,26}        30                                       $x:K p$                                        $\neg I$ : 27,29  
 {1}             31                                       $x:p \rightarrow K p$                                        $\rightarrow I$ : 26,30

We have thus arrived at the conclusion that if all truths *can* be known, than all truths *are* known—Fitch’s paradox of knowability.

### 3.3 Philosophical implications of the paradox—an illustration

The purpose of this section is to demonstrate that Fitch’s paradox is not just of logical significance—it also makes a genuine and insightful philosophical contribution.

Timothy Williamson uses Fitch’s paradox in his book *Knowledge and its limits*<sup>7</sup> to demonstrate, predictably, what the limits of our knowledge are. Let us briefly examine how. Williamson labels the thesis that all truths are known as *strong verificationism* (SVFR):

$$SVFR: \forall p (p \rightarrow K p)$$

<sup>7</sup> (Williamson, T., 2000).

This is, as Williamson puts it, an “insane sounding thesis” (p. 271). The more plausible sounding thesis that all truths are knowable Williamson calls *weak verificationism* (WVER):

$$\text{WVER: } \forall p (p \rightarrow \diamond K p)$$

Obviously, the stronger thesis implies the weaker one, since all that is known can be known. But in order to demonstrate some limits to our knowledge, Williamson uses Fitch’s paradox to demonstrate that the converse also holds—WVER implies SWER. They are therefore equivalent, and any objection to the insane-sounding thesis will apply to the more plausible formulation as well.

#### 4 SUMMARY

We put forth two desirable qualities of natural deduction systems at the beginning of this paper. The first—that it provide the same theorems as an axiomatic theory by way of making and following arbitrary assumptions—has clearly been met: we have derived all the axioms needed to prove one famous theorem. Regarding the second property, we have used the system to model the principles of reasoning present in the informal proof. So this was, in a manner of speaking, a textbook example of what natural deduction is supposed to do. Moreover, the formal principles come with no added complexity for someone familiar with first-order natural deduction, yet they are able to contribute to a fruitful philosophical debate.

#### APPENDIX

##### *Example 2.1*

{1}	1	$x: \Box (A \wedge B)$	$P$
{2}	2	$xRy$	$P^*$
{1,2}	3	$y: A \wedge B$	$\Box E: 1,2$
{1,2}	4	$y:A$	$\wedge E: 3$
{1}	5	$x: \Box A$	$\Box I: 2,4$
{6}	6	$xRz$	$P^*$
{1,6}	7	$z: A \wedge B$	$\Box E: 1,6$
{1,6}	8	$z:B$	$\wedge E: 7$
{1}	9	$x: B$	$\Box I: 6,8$
{1}	10	$x: \Box A \wedge \Box B$	$\wedge I: 5,9$

*Example 2.2*

{1}	1	$x: \Diamond (A \wedge B)$	$P$
{2}	2	$y: A \wedge B$	$P^*$
{3}	3	$xRy$	$P^*$
{2}	4	$y:A$	$\wedge E: 2$
{2,3}	5	$x: \Diamond A$	$\Diamond I: 3,4$
{1}	6	$x: \Diamond A$	$\Diamond E: 1,2,3,5$
{7}	7	$z: A \wedge B$	$P^*$
{8}	8	$xRz$	$P^*$
{7}	9	$z:B$	$\wedge E: 7$
{7,8}	10	$x: \Diamond B$	$\Diamond I: 8,9$
{1}	11	$x: \Diamond B$	$\Diamond E: 1,7,8,10$
{1}	12	$x: \Diamond A \wedge \Diamond B$	$\wedge I: 6,11$

REFERENCES

Barwise, J., Etchemendy, J., 2003, *Language, Proof and Logic*. Stanford, CSLI.

Basin, D., Matthews, S., Vigano, L., 1998, Natural deduction for non-classical logics. In *Studia Logica* 60, 119-160.

Brogaard, B., Salerno, S., 2008, Fitch's paradox of knowability. In E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*. Url=<(http://plato.stanford.edu)>.

Fitch, F., 1963, A logical analysis of some value concepts. *The Journal of Symbolic Logic* 28(2), 135-142.

Pelletier, F., 2000, A history of natural deduction and elementary logic textbooks. In Woods, J., Brown, B. (ed.), *Logical Consequence: Rival Approaches* vol 1. 105-138. Oxford, Hermes.

Williamson, T., 2000, *Knowledge and its Limits*. Oxford, Oxford UP.

JIRÍ RAČLAVSKÝ

## Partiality and Tichý's Transparent Intensional Logic

**Abstract.** The paper focuses on treating partiality within Tichý's logical system. Tichý's logic is two-valued and type-theoretic. His simple theory of types (and the deduction system for it) accepts both total and partial functions. Tichý's late framework is explicitly ramified. So-called constructions (roughly: algorithms) construct, e.g., values of functions at arguments; in some cases, however, they do not construct anything at all. This special partiality phenomenon is discussed in the second part of the paper.

### 1 INTRODUCTION

It will be convenient to begin with a sketch of the history of Pavel Tichý's transparent intensional logic (briefly *TIL*). One can find roots of TIL already in 1960s when Tichý construed intensions (functions from possible worlds) as classes of algorithms-procedures which were considered as meanings of (empirical) expressions; see *Intensions in Terms of Turing Machines* (Tichý 1969) reprinted in (Tichý 2004; hereafter *CP*). In the very beginning of the 1970's, Tichý modified Church's typed lambda calculus—accepting not only individuals and truth-values but also possible worlds; see *An Approach to Intensional Analysis* (1971) in *CP*. The system differs significantly from that developed by Montague (and Montagovians); the lack of space prevents me to give here a comparison (*cf.* Tichý's own remarks in *CP* 132-137 and the paper *Two kinds of intensional logic*, *CP* 307-325).

A new era of TIL (now explicitly designated by this name) is embodied in the large monograph (Tichý 1976) which remained unpublished. In this book,  $\lambda$ -terms and constructions recorded by them are explicitly distinguished (we will return to this issue later). Secondly, partial functions are admitted. Thirdly, natural deduction for the system is exposed. Selected parts of the book were



published as papers in the second half of 1970s; an exception is (Tichý 1982), which is a condensed paper on deduction.

In 1978 (*cf.* CP 269-270), Tichý added the temporal parameter; intensions are thus considered as functions from (possible worlds, time-moment) couples. Interesting logical analyses of temporal discourse (tenses, etc.) and episodic verbs were published by Tichý in 1980 (see CP). A more important modification of TIL is suggested in (Tichý 1988); Tichý exposed there a remarkable type theory which combines, in fact, simple and ramified type theory.

## 2 ADOPTING PARTIALITY

Let us begin with the recognition that there are both total and partial functions. Since many phenomena are to be modelled by partial functions (e.g., the chronology of American presidents in the actual world, or the individual concept “the king of France”), it is natural to accept them.<sup>1</sup>

Sometimes it is held that three-valued logic (3V-logic) captures partiality and so it is identical with two-valued logic which adopts partiality (2VP-logic). This is, however, a questionable matter. For 3V-logic—recognizing T (true), F (false), and U (unknown, undecided, ...)—is a logic with total functions only. On the other hand, 2V-logic recognizes T and F and accepts also a lack of a value for some function(s). Thus domains of truth-values of 3V-logic and 2VP-logic do differ.<sup>2</sup> For instance, there are 27 unary 3V-truth-functions but there are just 9 total and partial unary 2VP-truth-functions:

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$	$f_9$
T	T	T	T	F	F	F			
F	T	F		T	F		T	F	

Clearly, the function  $f_4$  is classical negation (often denoted by ‘ $\neg$ ’); it is, however, entirely missing in 3V-logic (this is why we should say that 3V-logic is not a classical logic). Of course, the 3V-function  $T \rightarrow F$ ,  $F \rightarrow T$ ,  $U \rightarrow U$  looks like a counterpart of  $\neg$ . For an obvious reason, however, plenty of 3V-functions cannot be counterparts of any 2VP-functions. (I will return to the problem of representation at the end of the paper.)

Once partial functions are admitted, strange phenomena appear. For instance, Schönfinkel’s reduction does not work because one multi-argument

<sup>1</sup>The reader knows that Imre Ruzsa stressed the importance of partiality (*cf.*, e.g., 1.2 in Ruzsa 1991).

<sup>2</sup>Of course, the acceptance of U or “gap” (as we may call it) is governed by the same intuition.

( $m$ -ary;  $m > 1$ ) partial function corresponds to more than one 1-argument function (Tichý 1982, 59-60); thus multi-argument functions are irreducible entities. Before we proceed further, let me introduce some notions.

### 3 TICHÝ'S SIMPLE THEORY OF TYPES

Tichý's simple theory of types—e.g., (Tichý 1982, 60)—treats both total and partial functions. It is quite general, since it has an unspecified basis  $B$ :

Let  $B$  consist of mutually non-overlapping collections of objects.

a) Any member of  $B$  is a *type over B*.

b) If  $\zeta, \zeta_1, \dots, \zeta_m$  are (not necessarily distinct) types over  $B$ , then  $(\zeta\zeta_1\dots\zeta_m)$ , which is a collection of all total and partial functions from  $\zeta_1, \dots, \zeta_m$  into  $\zeta$ , is a *type over B*.

The (specific) basis of TIL comprises  $\iota$  (individuals),  $o$  (truth-values T and F),  $\omega$  (possible worlds), and  $\tau$  (time-moments/real numbers). *Intensions* are functions from  $\omega$  to (total or partial) chronologies of  $\zeta$ -objects (a chronology is a function of type  $(\zeta\tau)$ ). Briefly speaking, intensions are functions from (possible world, time-moment) couples. ' $((\zeta\tau)\omega)$ ' will be abbreviated to ' $\zeta_{\tau\omega}$ '. *Propositions* are of type  $o_{\tau\omega}$ ; *properties of individuals* are of type  $(o\iota)_{\tau\omega}$ ; individual offices (Tichý's term) are of type  $\iota_{\tau\omega}$ ; etc. Objects which are not intensions may be called extensions. For instance, classical unary ( $\neg$ ) or binary ( $\wedge, \vee, \rightarrow, \leftrightarrow$ ) truth-functions are of types  $(oo)$  and  $(ooo)$ , respectively; classical quantifiers ( $\forall^\zeta, \exists^\zeta$ ) are of type  $(o(o\zeta))$ ;  $=^\zeta$  is of type  $(o\zeta\zeta)$  ( $^\zeta$  will be suppressed).<sup>3</sup>

### 4 CONSTRUCTIONS

To introduce the idea of constructions, consider the function:

$$\begin{aligned} 1 &\rightarrow -2 \\ 2 &\rightarrow 1 \\ 3 &\rightarrow 6 \\ &\vdots \\ &\vdots \end{aligned}$$

This function can be reached by (infinitely) many different (mathematical) procedures. For instance, it is induced by multiplying an integer by itself and subtracting three from the result (i.e. by  $(n \times n) - 3$ ) or by adding an integer to its square and subtracting what one gets by adding three to the integer from

<sup>3</sup> Ruzsa's type theory (cf. Ruzsa 1989, 3) does not allow some (types of) intensions which are admitted by Montagovians and Tichý. It should be added here that Tichý accepts not only functions of type  $\zeta_{\tau\omega}$  but also of type  $\zeta_\tau$  or  $\zeta_\omega$  (such functions are not called intensions in the present text).

the result (i.e. by  $(n^2 + n) - (n + 3)$ ). To every such intuitive procedure, there corresponds a certain Tichý (numerical) construction. Tichý used  $\lambda$ -terms to record constructions and one may view constructions as so-called intensional (i.e. not extensional) senses of  $\lambda$ -terms. To get another analogy, recall hyperintensions (“structured meanings”) often urged within the logical analysis of natural language. It seems also that Frege’s *Sinn* or Russell’s (structured) propositional functions are predecessors of constructions.

Unfortunately, a rigorous definition of constructions cannot be expounded here, see (Tichý 1988, 56-65) for that purpose. Omitting here so-called single and double execution, there are four kinds of constructions; their brief characterization is as follows. Let  $X$  be any object (a construction or non-construction) and  $C$  any construction; let  $v$  be any valuation (it is a field that consists of sequences of objects of given types):

1. *Trivialization*  ${}^0X$   $v$ -constructs  $X$  (i.e.  ${}^0X$  takes  $X$  and leave it as it is).
2. *Variable*  $x_k$   $v$ -constructs the  $k$ th member of the sequence of objects of a given type.
3. *Composition*  $[CC_1\dots C_m]$   $v$ -constructs the value of the function constructed by  $C$  at the string of objects (i.e. the argument for that function) which are constructed by  $C_1, \dots, C_m$ ; if  $C$  or  $C_1$  (etc.) does not  $v$ -construct such object(s) or the function is undefined for that argument,  $[CC_1\dots C_m]$  is *v-improper*—it does not  $v$ -construct anything at all.<sup>4</sup>
4. *Closure*  $\lambda x[\dots x\dots]$   $v$ -constructs, in a nutshell, the function which takes particular values of  $x$  to the objects  $v$ -constructed by  $[\dots x\dots]$  on the respective valuations (e.g.,  $\lambda n[{}^{n^0 \times n}0-03]$   $v$ -constructs the function sketched above).

One may thus say that these four kinds of constructions are objectual correlates of constants, variables (as letters), applications, and abstractions of  $\lambda$ -calculi. Realize, however, that constructions are not expressions—they are language-independent entities (the proper subject of Tichý’s approach are constructions, not expressions of some formal language). For instance, the term ‘ $\lambda n[{}^{n^0 \times n}0-03]$ ’ denotes (stands for) the construction  $\lambda n[{}^{n^0 \times n}0-03]$ . Realize also that constructions are not set-theoretical entities. Note that the term ‘ $\lambda n[{}^{n^0 \times n}0-03]$ ’ denotes the procedure as such, not the aforementioned function constructed by  $\lambda n[{}^{n^0 \times n}0-03]$  (analogously, ‘ $[{}^{08^0 \div 0}2]$ ’ denotes the construction  $[{}^{08^0 \div 0}2]$ , not its result—the number 4).

<sup>4</sup>The usual argument for the adoption of hyperintensions is this. Intensional semanticist suggests that all true mathematical sentences denote one and the same proposition (which is true in all possible worlds). Consequently, ‘Xenia believes that  $3+4=7$ ’, ‘ $49 \div 7=7$ ’.: ‘Xenia believes that  $3+4=49 \div 7$ ’ is rendered as a valid inference which is obviously not. Hence more fine-grained entities than intensions are needed to be explications of meanings. For another reason consider ‘Xenia calculates  $3 \div 0$ ’; the sentence surely describes the agent as related to a certain calculation, not to its (non-existing) result.

For non-circularity conditions, Tichý introduced a ramified theory of types. Its definition in (Tichý 1988, 66) has three parts: (a) types of (“classical”) set-theoretic objects (*cf.* the simple-type theoretic part above), (b) types of constructions (some constructions are first-order constructions, belonging to the type  $*$ <sub>1</sub>, other constructions are second-, third-, ..., *n*-order constructions), (c) types of functions from/to constructions.

In the mid-1970s, Tichý already suggested that constructions are explications of (natural-language) meanings—having thus the following semantic scheme:

an expression *E*  
*expresses* (means) in *I*.  
 the construction, which is the *meaning* (or logical analysis) of *E* in *I*.  
*constructs*  
 an intension / non-intension / nothing (*cf.* ‘3=0’), which is the *denotatum* of *E* in *I*.

The value of an intension in a possible world *w*, time-moment *t* is the *referent* of an empirical expression *E* (such as ‘dog’, ‘the king of France’, ‘It rains in London’); the denotatum and referent of a non-empirical expression are understood as identical.

For example (let *w* and *t* be variables *v*-constructing possible worlds and time-moments, respectively):

‘The king of France is bald’	an expression <i>E</i>
$\lambda w \lambda t [{}^0\text{Bald}_{w,t} {}^0\text{KF}_{w,t}]$	the construction expressed by <i>E</i> <sup>5</sup>
$\langle w_1, t_1 \rangle \rightarrow \text{T}$	the proposition denoted by <i>E</i>
$\langle w_2, t_2 \rangle \rightarrow$	(i.e. gap)
$\langle w_3, t_3 \rangle \rightarrow \text{F}$	
etc.	
T	the referent of <i>E</i> in $w_1, t_1$

It is not difficult to show that this semantic theory is capable to deal with puzzles created by “intensional” and “hyperintensional” contexts.

## 5 PARTIALITY AND FAILURE OF CLASSICAL LAWS

From the objectual viewpoint, logical laws are not strings of letters but constructions. It is clear that (let *o* be a variable *v*-constructing truth-values):

$$[{}^0\forall \lambda o [o^0 \vee [{}^0 \neg o]]]$$

<sup>5</sup> Trivializations of well-known mathematical or logical functions will be written in the infix manner (e.g., ‘[ ${}^0 8^0 \div 2$ ]’ instead of ‘[ ${}^0 \div 8^0 2$ ]’).

is tautological (the variable  $o$  is always a  $v$ -proper construction). However, this law is scarcely remarkable, one would rather declare that for any proposition, it obtains in  $w$ ,  $t$  or it does not obtain in  $w$ ,  $t$  (the excluded middle). Let  $p$  be a variable  $v$ -constructing propositions (i.e. objects of type  $o_{\tau_w}$ ). Then the following construction (which can be closed by  $[{}^0\forall\lambda w[{}^0\forall[\lambda x;$  similarly below) is contradictory:

$$[{}^0\forall\lambda p[p_{w,t}{}^0\vee[{}^0\neg p_{w,t}]]]$$

Since if some proposition is undefined in  $w$ ,  $t$ , then  $\lambda p[p_{w,t}{}^0\vee[{}^0\neg p_{w,t}]]$   $v$ -constructs a partial class (partial characteristic function) which is empty, thus  $\forall$  takes it to the truth-value F.

We get an analogous failure for the carelessly formulated De Morgan law for exchange of quantifiers. Let  $P$  be any construction of a proposition where  $P$  contains  $x$  as its free variable (e.g.,  $\lambda w\lambda t[x^0=^0KF_{w,t}]$ ):

$$[[{}^0\neg[{}^0\exists\lambda xP_{w,t}]]]{}^0\leftrightarrow[{}^0\forall\lambda x[{}^0\neg P_{w,t}]]$$

The construction  $\lambda w\lambda t[x^0=^0KF_{w,t}]$  (which is reducible to  $\lambda x[x^0=^0KF_{w,t}]$ ) can  $v$ -construct a partial class which is empty thus  $\forall$  takes it to F, not to T as we wish.

To avoid the destructive power of partiality, formulating thus the correct versions of the laws, I suggest utilizing a “totalizer” overcoming the trouble. In Tichý’s framework, there are three kinds of properties “be true” due to their applicability to (a) propositions, (b) constructions, (c) expressions (relatively to a given language  $L$ ). Each kind has several variants; the (a)-kind has only two. The “*partial*” *truth property* of propositions (i.e. an object of type  $(o_{\tau_w})_{\tau_w}$ ) can be defined as:<sup>6</sup>

$$[{}^0\text{True}^{\pi P}_{w,t}p] \equiv p_{w,t}$$

thus certain propositions are not in the extension or the anti-extension of that property (in  $w$ ,  $t$ ). The “*total*” *truth property* of propositions can be defined as:

$$[{}^0\text{True}^{\pi T}_{w,t}p] \equiv [{}^0\exists\lambda o[[p_{w,t}{}^0=o]{}^0\wedge[o^0=^0T]]]$$

A partial proposition having no value in  $w$ ,  $t$  belongs to the anti-extension of the property—it is not true in  $w$ ,  $t$ .

Using  ${}^0\text{True}^{\pi T}$  for “totalizing”, the correct law of excluded middle is:

<sup>6</sup> ‘ $C_{w,t}$ ’ abbreviates ‘ $[[C_{w,t}]]$ ’. Of course,  ${}^0KF$  is a simplification. The procedure consists in taking (a) the property “popular”, (b) applying it to  $w$  and  $t$  (values of  $w$  and  $t$ ), getting thus the extension of “popular”, and then (c) taking “the king of France”, (d) applying it to  $w$  and  $t$ , getting thus the individual who fills that office, and (e) asking whether that individual (if any) is in that extension—yielding thus T or F (analogously for other  $w$ ’s and  $t$ ’s).

$$[{}^0\forall\lambda\rho[[{}^0\text{True}^{\pi^1}_{w'}\rho]{}^0\vee[{}^0\neg[{}^0\text{True}^{\pi^1}_{w'}\rho]]]]$$

and the correct De Morgan law is:

$$[[{}^0\neg[{}^0\exists\lambda xP_{w'}]]]{}^0\leftrightarrow[{}^0\forall\lambda x[{}^0\neg[{}^0\text{True}^{\pi^1}_{w'}P]]]$$

So it is clear that partiality affects the rules for substitutivity (e.g., whether  $\rho$  can be substituted by  $P_{w'}$ ), which lead Tichý to the sophisticated theory exposed in (Tichý 1982) and in Indiscernibility of Identicals (Tichý 1986), where he paid closer attention to constructions involving identity.

## 6 BETA-REDUCTION, ETA-REDUCTION AND PARTIALITY

But there are more complications with partiality—even classical  $\beta$ -reduction fails ( $\beta$ -reduction rule says that  $[\lambda x[\dots x\dots]C]$  is equivalent to  $[\dots C\dots]$ ). Consider:

1.  $\lambda w\lambda t[{}^0\neg[{}^0\text{True}^{\pi^1}_{w'}\lambda w\lambda t[{}^0\text{Bald}_{w'}\ {}^0\text{KF}_{w'}]]]$

(the analysis of ‘It is not true that the King of France is bald’)

2.  $\lambda w\lambda t[\lambda x[{}^0\neg[{}^0\text{True}^{\pi^1}_{w'}\lambda w\lambda t[{}^0\text{Bald}_{w'}x]]]{}^0\text{KF}_{w'}]$

(the analysis of ‘The King of France is such that it is not true that he is bald’)

The two constructions  $v$ -construct distinct propositions because 1.  $v$ -constructs a total proposition whereas 2.  $v$ -constructs a partial proposition (if there is no king of France in  $w$ ,  $t$ ,  ${}^0\text{KF}_{w'}$  is  $v$ -improper, so the proposition is gappy). Thus 2. is not  $\beta$ -reducible to 1.<sup>7</sup>

In (Tichý 1982, 67),  $\beta$ -reduction and  $\beta$ -expansion are explained as deduction rules (The Rule of Contraction/Expansion). Tichý’s rule of  $\beta$ -reduction contains an explicit condition that the construction  $C$ , which is substituted, is not  $v$ -improper. So conditioned,  $\beta$ -reduction preserves equivalence of constructions.<sup>8</sup>

Moreover,  $\eta$ -reduction (that  $\lambda x[Cx]$  is reducible-equivalent to  $C$ ) fails as well (Raclavský 2009, 283). Consider  $[{}^0Fy]$ , where  ${}^0F$   $v$ -constructs a function of type  $((\theta\zeta)\zeta)$  which is undefined for the  $\zeta$ -object assigned to  $y$  by  $v$ . Thus  $[{}^0Fy]$  is  $v$ -improper, it  $v$ -constructs nothing at all. But  $\lambda x[[{}^0Fy]x]$  does  $v$ -construct an object, namely a function of type  $(\theta\zeta)$  which is undefined for the  $\zeta$ -object assigned to  $x$  by  $v$ . Hence  $\lambda x[[{}^0Fy]x]$  cannot be equivalent to  $[{}^0Fy]$ . The remedy (*ibid.*) is the same as Tichý’s conditioning of  $\beta$ -reduction.

<sup>7</sup> Here  $\equiv$  means inter-derivability of two constructions (*cf.*  $\leftrightarrow$ , in CP 489); the two constructions flanking  $\equiv$   $v$ -construct one and the same object (a truth-value in this case) or they both  $v$ -construct nothing at all. The construction  $[{}^0\text{True}^{\text{mp}}_{w'}\rho]$  can be closed by  $\lambda w\lambda t[\lambda\rho]$  and then  $\eta$ -reduced to  ${}^0\text{True}^{\text{mp}}$  (which is used below). See (Raclavský 2008) for more.

<sup>8</sup> *Cf.* (Duží 2003) for more.

7 ANOTHER WAY OF REPAIRING PARTIALITY

When defining various concepts one sometimes needs to overcome partiality of some function. For the sake of illustration, imagine that you sum salaries of various people—including the king of France (“...+the salary of(KF)+...”). Since there is no king of France, “the salary of(KF)” returns no number. But you need a certain number (zero in this case) because you do not want the final sum (“...+...+...”) to be undermined by the “local” partiality failure.

In Tichý’s logic, the delivering of a “dummy value” (e.g. zero) can be easily managed in the following way (see Raclavský 2009, 243). Consider a partial function  $F$  from (type-theoretically appropriate)  $x$ ’s to (type-theoretically appropriate)  $y$ ’s. Since  $F$  is partial,  $[^0Fx]$  can  $v$ -construct nothing; in such case, however, you need something—a dummy value. I suggest replacing  $[^0Fx]$  in a construction  $C$  (which is affected by partiality of  $F$ ) by the following construction which fulfils our demand:

$$\begin{array}{ll}
 [^0\text{Sng}\lambda z\{^0\text{If\_} & [^0\exists\lambda y\{y^0=[^0Fx]\}] \\
 \_ \text{Then\_} & [^0\exists\lambda o\{[o^0=[z^0=[^0Fx]]\}^0\wedge[o^0=^0T]\}] \\
 \_ \text{Else} & [z^0=^0\text{DummyValue}]\}]
 \end{array}$$

(the singularization function,  $\text{Sng}^\xi$ , takes one-membered classes of  $\xi$ -objects to their sole members, it is undefined otherwise;  $\text{if\_then\_else}$  is the well-known ternary truth-function, its trivialization is written in parts; note that there is an analogue of “it is true $^{\pi 1}$  that  $y=F(x)$ ” in the second line).

8 THREE-VALUED FUNCTIONS REPRESENTED BY PROCEDURES

To conclude this short paper, 2VP-logic incorporating procedures (constructions) is capable to capture the intuition which underlies 3V-logic, if 3V-functions are modelled not by (partial) functions but by procedures. For instance, the 3V-function  $T \rightarrow F, F \rightarrow T, U \rightarrow U$  can be modelled by  $[^0\neg p_{ev}]$  because this construction behaves in an analogous way as that 3V-function: it returns  $T$  when the proposition  $p$  has (in  $w, t$ ) the truth-value  $F$  (and *vice versa*) but it returns nothing if the proposition  $p$  is undefined (in  $w, t$ ). To define procedures representing other 3V-functions is usually more involved—one must utilize  $^0\text{True}^{\pi 1}$ , often together with the dummy-value construction.

## REFERENCES

- Duží, M., 2003, Do we have to deal with partiality? *Miscellanea Logica* V, 45-76.
- Raclavský, J., 2008, Explications of kinds of being true (in Czech). *SPFFBU* B 53(1), 89-99.
- Raclavský, J., 2009, *Names and Descriptions: Logico-Semantical Investigations* (in Czech). Olomouc, Nakladatelství Olomouc.
- Ruzsa, I., 1991, *Intensional Logic Revisited*. Budapest, published by the author.
- Tichý, P., 1971, An approach to intensional analysis. *Noûs* 5(3), 273-297.<sup>9</sup>
- Tichý, P., 1976, *Introduction to Intensional Logic*. Unpublished book manuscript.
- Tichý, P., 1982, The foundations of partial type theory. *Reports on Mathematical Logic* 14, 57-72.
- Tichý, P., 1986, Indiscernibility of identicals. *Studia Logica* 45(3), 257-273.
- Tichý, P., 1988, *The Foundations of Frege's Logic*. Berlin, Walter de Gruyter.
- Tichý, P., 2004, *Pavel Tichý's Collected Papers in Logic and Philosophy*. V. Svoboda, B. Jespersen, C. Cheyne (eds.), Dunedin, University of Otago Publisher / Prague, Filosofia.

<sup>9</sup> It seems that 2.4.2.5 in (Ruzsa 1989) captures an analogous restriction of  $\beta$ -reduction.



## Prior on Radical Coming-into-Being

**Abstract.** Chapter VIII of *Papers on Time and Tense* (1968) on ‘Identifiable Individuals’ is the locus where Prior elaborates his polemic on whether radical coming-into-being is a *genuine de re possibility* of the individual substances. He argues for the conclusion that there is no such possibility concerning radical coming-into-being as opposed to piecemeal change. Therefore the property or predicate about an individual’s origin does not yield *de re* modality.

Prior considers the thought experiment of swapping the properties of two individuals through worlds. Say, we have in the actual world the person Julius Caesar with the usual Caesar-like properties. We have also Mark Antony as an inhabitant of the actual world with his own properties. Suppose that through chains of accessible worlds we reach a world such that Caesar has all and only the properties of Mark Antony and vice versa. Surely, this world is qualitatively indistinguishable from the original world. The thought experiment is typically explored for highlighting our intuitions whether the world resulting in complete property swap is different in any sense from the original world. Those who say that the two worlds are actually one, for all that matters only is what properties are instantiated and co-instantiated in a world and the instantiating entities are nothing over and above the properties they instantiate—are Leibnizians applying the criterion of L-indiscernibility. Others, who feel that there is a genuine difference between the two worlds, claim that Caesar and Antony still differ by their haecceistic properties of *being Caesar* and *being Antony*, respectively, whatever other properties of theirs are swapped. The usual retort from the Leibnizian camp is that haecceistic properties are not genuine properties.

The dispute about the identity condition of worlds and individuals cannot be settled in a short talk. What is interesting is how Prior finds his way out of the dilemma. He is neither Leibnizian nor haecceist. Moreover, he is not making the Kripkean point about the necessity of origin. Instead, he finds that the property which is necessarily exempt from property-swap is the property of originating from the actual ancestors one has. This property is resistant

to property exchange. If this is so, it helps to solve the problem of Leibniz-indiscernible worlds. And it has the virtue of not being a suspect haecceistic property, like that of being identical to Caesar; it is similar to it though in being an impure relational property.

After considering several property-swaps, Prior asks: 'can we not go further and suppose Caesar to have had the whole of Antony's life, including being born to Antony's parents?' (1968, 85.) The question is not about logical possibility since the proposition asserting Caesar's having born to Antony's parents is not inconsistent.

By contrast, construed as a question about a *temporal* possibility, we have a more substantial issue at hand. For now we can ask: '*when* was it possible?' And it is easy to see that '*after* his birth ... it was clearly *too late* for him to have had different parents.' (ibid.)

This insight is fairly obvious. Why not ascribe, then, the *de re* possibility of having different parents '*before* Caesar existed'? Intuitively speaking, the remoter the present is, viewed from a distant future, the more possibilities are still open concerning that present (taken indexically). Alas, the crucial point, highlighted by Prior, is that however broad the general possibility is in this case, 'there would seem to have been no individual identifiable as Caesar, i.e. the Caesar who we are now discussing, who could have been the subject of this possibility.' (ibid.)

Now Prior's point can be generalized to the possibility of actual individuals, too. My contribution to the problem consists in this suggestion: let us extend Prior's point to the *de re* temporal possibility of the coming into being of *actual* individuals thereby bringing forth its full metaphysical significance. The generalization is this:

If Caesar (or any other actual individual) could not have been the subject, before his birth, of the (later) unrealized possibility, equally, *he* could not have been the subject of the *later realized* possibility either. This means that none of us who was going to be born could have been the subject of a *de re* possibility of being (going to be) born – *i.e.*, at least not before our conception. And this amounts to saying that what is once actual is always preceded by what is non-possible, contradicting thus the logic of propositional modalities. According to these modalities the actual is never preceded by something impossible. "Precedence" is taken here in the sense of yielding existential conditions through temporal priority.

So we have to rule out not only the unrealized *de re* possibilities concerning origins. The realized courses of originating by birth are not possible either, at least *not* in the *de re* sense. The reason being that there is no identifiable individual to ascribe the *de re* possibility to. Hence, there is no identifiable individual to ascribe the putative *de re* necessity of origin either. The realized and the unrealized *de re* possibilities/necessities are in a symmetric situation

relative to each other: both require the semantic precondition that the term in a referential position successfully refer, which is not fulfilled in either case. Let me note that cases of origin as temporal *de re* modalities are treated like cases of empty names in extensional contexts.

Prior rightly claims that any genuine *de re* possibility/necessity presupposes that the subject of the modal ascription, even if just a 'thin' individual, be fixed referentially. Since this is not satisfied in the case of the putative *de re* possibility of origin, the latter cannot be regarded as a case of *genuine de re* possibility. Further, I suspect that Prior is doing more than simply calling attention to the lack of a demonstrative reference when there is not yet anybody before birth to refer to. I take it that Prior is emphasizing the absence of a *descriptive* device of referring to individuals to consider *de re* possibilities. As Prior notes 'as new distinguishable individuals come into being, there is ... a multiplication of distinguishable logical possibilities' (*op.cit.*, p.91.) Now it is fairly obvious that the basis of the distinction among individuals and their possibilities are of a qualitative nature. For example, there must be people identifiable qualitatively as different applicants to a job for opening the logical possibility of competition for that job. So I take it that Prior is implying some descriptive notion of individuals. It is a further question how 'thin' or 'thick' his implied notion of individuals is.

We have seen that the putative *de re* possibility of origin lacks the semantic precondition of reference. But then it is hard to see how one can satisfy the natural intuition that it is still meaningful to talk about the possibility of someone's having had a different origin.

The suggested solution is this: there is no subject, strictly speaking, of such *de re* possibilities. Our modal claims can be entertained only in the way that can be illustrated by Burleigh's example of a promise of giving someone a horse. Such promise may be disambiguated *not* as the promise of a specific horse but as the promise that can be fulfilled by giving someone *any* horse. So is the case with the possibility of origin. As Prior puts it, 'the possibility that *an* individual should begin to exist ... is like a promise of the second kind'. (*op.cit.*, p.86.) In other words, this possibility is 'general' rather than specific. This is tantamount to saying that the possibility of origin is not a *de re* but a *de dicto* possibility. And this is the solution to the seemingly paradoxical situation. It is possible *that* someone be born to such and such parents, but it is not possible *of someone* that *he* should be born to these or other parents.

So, Prior's solution to the problem of *de re* possibility with radical coming-into-being consists in denying that there is such *de re* possibility and satisfying, instead, our modal intuition with the *de dicto* form of possibility. *De re* possibility with respect to radical coming-into-being is, at best, a *post factum* possibility: as Thomas Aquinas put it, quoted by Prior, it is an 'accident' which 'is subsequent to the thing' that has already come into existence.

The ramification of the topic worth considering is the issue of temporal vs logical possibilities. The temporal possibility just discussed is obviously different from possibility in the logical sense. In the logical sense of 'could' Caesar could have been born to persons who actually turned out to be the parents of Mark Antony for there is no inconsistency in the proposition stating this course of events. (Leibnizians would, perhaps, object that there is, for the 'complete concept' of Caesar excludes the relational property of being born to those other people.) Logical possibility seems then to be permissive about a case explicitly ruled out by temporal possibility.

The gap between the two kinds of possibilities is smaller though: for, both possibilities as *de re* possibilities have an existential precondition such that 'before [Caesar] existed it was not logically or in any other way possible that he should *come to have* those people, or any other people, as his parents'. (*op.cit.*, p.92.) As possibilities dependent on an (empirical) existential precondition, the logical and the temporal readings of possibility behave much the same way.

Another ramification would be to compare radical coming-into-being with piecemeal change. It is clear from the foregoing that radical coming-into-being, in Prior's view, is not change in the piecemeal sense of an existing thing acquiring/dropping properties over time. Coming-into-being should not be taken, as he points out, to mean that 'once *X*'s non-being was the case and now its being is'. It should be taken to mean, instead, that 'it is *not* the case that *X was*, but it *is* the case that *X is*', and this does not express a change but two contrasting present facts'. (*op.cit.*, p.88.) Clearly, radical coming-into-being does not have the features of a genuine change. Radical coming-into-being is not unique, however, in this respect: Cambridge changes are typically not taken to be genuine changes either. Elsewhere I have discussed them.

Conclusion: the property of origin should be exempt from the range of properties affording *de re* locutions. This logical insight is backed by the metaphysical insight that radical coming-into-being does not constitute genuine change. The common source of these insights is that the property of origin (unlike other properties of individuals) lacks the existential precondition. *After* this precondition being fulfilled, the property of origin is an accident subsequent to the individual. As a consequence, the property swaps between individuals through worlds must stop at the impure relational property of origin.

The broader metaphysical moral is this: we have seen the limits of *de re* talk drawn by the property of origin or radical coming-into-being. It is all too familiar that *de re* locutions are found suspect from Quine on, for the weird metaphysics one gets when one quantifies in modal contexts. The present talk is obviously not a follow up on Quine's discontent with the 'invidious attitude' (as he calls it)

of Aristotelian essentialism.<sup>1</sup> Rather, by drawing the lines of applicability one can assess the merits of the metaphysical doctrines. As to the significance of Prior's argument, we have seen that it offers a genuine novel suggestion to the polemic about property-indiscernible worlds. He does not seek haecceistic differences in order to explain the difference of Leibniz-indiscernible worlds. Rather, Prior shows that there is a further difference which is responsible for the difference between any two qualitatively indiscernible worlds: it is the *difference of origin* of the individual inhabitants of the worlds. To repeat, the *property of origin* is neither purely qualitative, nor haecceistic, but a peculiar accident that can be had only *post factum* and it is resistant to property-swap.

#### REFERENCES

- Prior, Arthur, 1968, *Papers on Time and Tense*. Oxford, Clarendon Press.  
Quine, W.O, 1953/1963, Reference and Modality. In id., *From a Logical Point of View*. Cambridge, Mass., Harvard University Press.

<sup>1</sup> Quine (1953) characterizes 'Aristotelian essentialism' by its 'adopting an invidious attitude towards certain ways of uniquely specifying' an object and seeing these ways 'as somehow better revealing [the object's] "essence"' than alternative specifications.

## Analogy in Semantics

**Abstract.** The principle of compositionality seems too trivial and too restrictive at the same time. I propose that this principle should be seen as (part of) a definition of what “meaning” is in a theory that posits a very abstract concept of “meaning”, one very far from empirically testable reality. The principle of compositionality presupposes an analytic/atomistic approach to meaning, where the properties of a whole must be explained in terms of those of its constituent parts. In my paper, I will propose a *holistic* approach instead: I will emphasise the global features of signs. I will introduce the principle of *generalized compositionality*, which is based on the concept of similarities between forms and meanings. My generalized compositionality principle states that we interpret and produce complex signs by analogy, relying on our earlier experience on similar complex signs and their interpretation. This move, I believe, is necessary for moving towards a cognitively more realistic model, with a view to predicting frequency effects and other psychological factors of interpretation. As a side-effect, the dubious distinction between “literal” and “non-literal” interpretations disappears.

### INTRODUCTION

In this paper I will argue for a model of natural-language semantics that diverges in important ways from the main-stream approach, but it fits well within a more general cognitive framework in which language is viewed as a system of habits, in much the same sense as it was viewed in the late 1800s and in European structuralism.

I will start from two different families of problems, namely, the distinction between “literal” meanings and actual interpretations, e.g., in the case of metonymical language use (section 1.1), and an ambiguity related to disjunctions used in negative contexts (section 1.2). I will show that both of these types of

problems pose serious challenges for the traditional machinery of semantics that posits a compositional “translation” process that produces logical formulae.

I then turn to the formulation of and the inherent problems with the principle of compositionality (section 2.1), and I will propose a different principle, which I call *generalized compositionality*, which avoids the pitfalls of traditional compositionality (section 2.2). I will argue that this alternative formulation can serve as the fundamental principle of *analogy-based semantics*, a view of semantics that does not require “underlying”, abstract syntactic representations or a distinction between “literal” and “non-literal” interpretation. I will briefly return to the linguistic problems that constitute the starting-point of the discussion, and outline ways of approaching them in analogy-based semantics. I conclude, in section 2.6, by sketching a model of associative memory that can serve as an implementation of the machinery of analogy-based linguistic theory.

## 1 TWO PROBLEMS IN LINGUISTIC SEMANTICS

### 1.1 *Non-literal meaning and abduction*

As pointed out in (Hobbs & al. 1993), even the simplest sentences, like his *The Boston office called*, contain *implicit information*, i.e., propositions that are *plausible* for the audience to assume, and without which we cannot speak of making sense of, interpreting, or understanding what the sentence “literally” says. In this sentence, for example, the audience has to figure out (among other things) that *the Boston office* is an office located in Boston, and that someone working at that office (rather than the office itself) placed the call. (Hobbs & al. 1993) propose to consider this mechanism a case of abduction, i.e., to consider the “literal” interpretation of the sentence as a conclusion, and the plausible assumptions the audience has to take on as missing premises, which would then make the conclusion true.

Unfortunately, the two key terms that I have just employed, *plausibility* and *literalness* are very hard to grasp, and there is hardly a consensus surrounding them. For example, it is not clear why ‘someone working at the Boston office called’ is more plausible than, say, ‘someone who happened to be at the Boston office called’; or, to take another example, it is not clear what the “literal meaning” of a compound like *Boston office* is: ‘an office somehow related to Boston’ (as (Hobbs & al. 1993) assumes, in which case we have to rely on abduction again to get the metonymical interpretation) or ‘an office located in Boston’, and it is also not clear whether this expression is ambiguous, i.e., whether it has more than one “literal meaning”.

From the linguistic point of view, the approach taken in (Hobbs & al. 1993) illustrates nicely the paradigm going back at least to (Grice 1967), according to which utterances can be assigned context-independent, purely linguistic,

“literal” *meanings*, which leave much uncertainty as to the appropriate interpretation in a given context. In Grice’s (1967) view, it is the anomalies and contradictions arising from this basic “meaning” that give rise to all sorts of implicit information that the audience tends to assume, in order to resolve the anomalies, and thereby arrive at a contextually determined “non-literal” interpretation. The process often involves ontological (real-world) knowledge and pragmatic considerations. In (Hobbs & al. 1993), no anomaly is needed to trigger this process: implicit information is always retrieved until a *maximal consistent* stock of background knowledge is incorporated into the interpretation through abduction (by assuming only *maximally plausible* pieces of information).

This model of understanding heavily relies on a dichotomy between “linguistic knowledge” (driving the translation mechanism that produces a “meaning” from a linguistic expression), on the one hand, and “background knowledge”, from which plausible assumptions are retrieved by the abduction process. But this dichotomy is disputable. For example, in the case of *Boston office*, the abduction process must use an axiom in the knowledge base according to which “located\_at( $y, x$ )  $\rightarrow$  rel( $x, y$ )”, where the translation of *Boston office* contains “office( $x$ )  $\wedge$  boston( $y$ )  $\wedge$  rel( $x, y$ )”, so “ $x$  is located at  $y$ ” is a plausible premise for  $x$  standing in the relation “rel” with  $y$ , and the conclusion is part of the “literal meaning” of *Boston office*. But it is not at all clear why just the use of the mysterious relation “rel” is part of the “linguistic” module, whereas possible interpretations of such compounds are located in the “background knowledge”, although they are clearly of a linguistic character, too (as shown by the fact that the same type of compounds cannot be interpreted in this way in Hungarian: there, a suffix must be added to *Boston* in order for it to refer to the location of the office).

In general, there is no universal recipe for deciding whether a certain ingredient of understanding (in our example, ‘the office is located in Boston’) should stem from “linguistic knowledge” (in our example, this would amount to a particular “compounding rule” to this effect), or from “background knowledge” (in our example, this corresponds to the approach in Hobbs & al. 1993). There is no accepted philosophical, linguistic or psychological evidence for deciding either way for any particular case.

The shakiness of the distinction between “linguistic” and “background” knowledge is not a purely theoretical problem. In particular, it is challenging for any potential *mental model* of the mechanism of understanding, in particular a model of *acquisition*, on the one hand, and a model capable of predicting *performance data*, on the other. It is not obvious, to say the least, how the distinction between “linguistic” and “background” knowledge could be learned, and I do not believe there is any psycholinguistic evidence to the effect that the “translation” of an expression such as *the Boston office* (i.e., retrieving its



“literal meaning”) should be separable from the process of interpreting it using background knowledge.

## 1.2 Disjunction of NP's

Let me now turn to a completely different problem, or rather a family of problems, unrelated at first sight to those explained in 1.1. These problems, which have puzzled semanticists for a long time (e.g., Szabolcsi 2002; Szabolcsi & al. 2004), are connected to the interpretation of *disjunction* in various embedded positions. For example, consider the following ambiguity:

- (1) *He didn't close the door or the window.*
- a. 'He left both the door and the window open'
  - b. 'He left either the door or the window (or both) open'

Szabolcsi (2002) explains the ambiguity in an essentially syntactic way (although she relies on features that can be viewed as having semantic content). She claims that, under one interpretation, negation “has a wider scope than” disjunction, whereas the opposite holds when the other reading is to be obtained:

- (1') a.  $\neg(\text{'closed the door'} \vee \text{'closed the window'})$   
 b.  $(\neg\text{'closed the door'} \vee \neg\text{'closed the window'})$

Note that in order to obtain these readings, highly unnatural (abstract, covert) syntactic structures have to be assumed: both of these formulae are very distant from the surface structure of the sentence in (1). Thus, this mechanism requires a non-deterministic “translation” step from any theory that respects compositionality.

The situation is even worse when negation is not explicit, but inherent in the predicate, as in the following example:

- (2) *You forgot to close the door or the window.*
- a.  $\neg(\text{'remembered to close the door'} \vee \text{'remembered to close the window'})$
  - b.  $(\neg\text{'remembered to close the door'} \vee \neg\text{'remembered to close the window'})$

In order to get the interaction of negation and disjunction here, one has to assume a representation in which the negation inherent in *forget* is made explicit by converting it into *not remember*.

While the problem explained in 1.1 allegedly concerns how we arrive at the desired interpretation from the “literal”, “linguistic” meaning, the problem of interpreting disjunctions seems related to the “translation” itself, i.e., to how we get from a surface utterance to the “literal”, “linguistic” meaning.

Under the standard approach, in order to account for ambiguities like the one presented here, we must assign two different logical formulae to one and the same utterance, depending on factors that are not entirely clear.

In both cases, there is a gap between a surface expression and its interpretation, i.e., both problems challenge the traditional, *compositional* approach to semantics, which consists in assigning a logical representation to utterances in a systematic way, then interpreting that representation with respect to a model of the world. It looks like we have to revise the concept of compositionality in the sense it is normally used while, for obvious reasons, we do not want to deny the systematic character of interpreting natural-language utterances.

## 2 GENERALIZING COMPOSITIONALITY

### 2.1 *Problems with compositionality*

The principle of compositionality is usually stated as follows:

(3) *The principle of compositionality*

The meaning of a complex expression is a function of the meanings of the constituents which syntactically constitute it.

This principle, attributed to Frege, is considered the starting-point of modern semantics, aiming at establishing and examining the systematic relationship between form and meaning in natural language. It is commonly thought of as saying something fundamental (in fact, almost trivial) about meanings and the science of meanings. On the other hand, it has puzzled most modern semanticists in various ways because of the multitude of phenomena that it does not naturally apply to. To mention a couple of those, there is the context dependence of the use of many, if not all, linguistic expressions (so the context of utterance can be seen as an additional factor, not mentioned in the principle of compositionality, potentially influencing the possible use of a complex expression), and there are legions of obviously complex expressions in all natural languages the meaning of which is “non-compositional” or idiomatic to a smaller or greater extent (which means that their meanings are more or less unrelated to the meanings of their constituent parts).

As for context dependence, the received view seems to be that the principle of compositionality is a methodological axiom that demarcates what is covered by part of the definition of meaning exactly by eliminating those properties of linguistic use that depend on contextual information. That is, context dependence does not contradict compositionality; to the contrary, compositionality carves out just those aspects of the use of linguistic signs that are independent

of context,<sup>1</sup> and constrains the term *meaning* by excluding all other aspects (and tacitly relegating them to some other field of study, say, “pragmatics”).

The problem of those expressions that are idiomatic to some extent is a much tougher one. We could approach this problem in the same vein as context dependence, by saying that the principle of compositionality limits semantic study to the so-called “compositional” (non-idiomatic) constructions, but it is not at all clear where the boundaries would lie, and what such a semantics could do with mixed cases when a complex expression has both idiomatic and non-idiomatic aspects. For example, an idiom like *spill the beans* has the same aspectual and thematic features as its main verb *spill* does (it is agentive and expresses an accomplishment or an achievement), so it is “transparent” in this respect. (As a matter of fact, the view that there is a concept of “meaning” that abstracts away all context dependence, can be challenged in a similar vein, but I will not go into that discussion here.)

So buying into the principle of compositionality seems to represent an important sacrifice for a truly scientific approach to meanings. On the other hand, it can be shown that it is ridiculously easy to still satisfy the principle, given the fact that the concepts of *meaning*, *syntactic constitution* and *function* figure in it without any further specification or constraint. The principle imposes no constraint on ambiguities (of the expressions constituting a complex one) and on the process on their resolution. For example, take the understanding of *The Boston office called*: the principle of compositionality does not constrain how *office* and/or *call* are to be disambiguated. For example, it is not a violation of compositionality if one uses “global” information on what the whole utterance (or even the text that it occurs in) is about, or what the speaker’s intentions are. This means that, effectively, there can be wildly non-compositional steps in a strictly speaking compositional interpretation process.

So both “non-compositional” constructions and context dependence can be “explained away” by appealing to the presence of multiple (possibly infinite) ambiguities and obscure mechanisms for their resolution. Meanings can be of any nature whatsoever, so it is possible to encode not only semantic, but also formal properties in them. As an extreme case, we can posit that the “meanings” of constituent expressions specify their interpretation for each (type of) complex expression they occur in, which is compatible with the definition, yet it would make compositionality entirely empty. As for the concept of “function”, a function can do just about anything and, by taking this to the extreme, one can arrive at wildly unnatural, yet compositional mechanisms of interpretation—like ignoring the meaning of a constituent: for other examples, see, e.g., (Zadrozny

<sup>1</sup>Sometimes context still plays a role, when reference to context is said to be built into “meanings”. For example, the fact that the personal pronoun *I* refers to the speaker is said to be part of its “meaning”, while the context determines who the speaker actually is. Cf. (Gendler Szabó 2000) for a summary.

1994)). Finally, the principle of compositionality contains no proviso to the effect that syntactic structure be determined independently of semantics. As a matter of fact, most syntactic constituency tests (such as substitutability or mutual information) are implicitly based on the greater semantic cohesion between certain structural elements than that between others within a construction. So there is ample space for “playing around” with syntactic structure in order to satisfy compositionality.

## 2.2 *Generalized compositionality*

As I have pointed out in the previous section, the principle of compositionality is both too strong and too weak, too restrictive and too liberal at the same time. I have also mentioned that, still, it is essential to adopt it in some form or another, inasmuch as it captures one of the leading ideas of the linguistics of the late 19th century, namely, that the productive aspect of natural languages is due to the *systematicity* of our ability to understand utterances. That is, we can only understand an utterance if it is built up in a similar way to utterances we have seen earlier, and we interpret it by *analogy* to how those earlier utterances were used. By the same token, we are able to communicate through utterances if we utter ones that are sufficiently similar to those the audience is familiar with, and we have to take into account that they will be interpreted in the same way as the audience’s earlier experiences dictate.

The key concepts here are *recurrence* and *similarity* rather than “meaning”, “function” and “constituency”, but the basic idea, I believe, is the same as in the case of the principle of compositionality. Not only elementary signs, but also their combinations must show a sufficient amount of similarity to combinations seen earlier in order for them to be understandable, which means that the way they are combined must also be recurrent, and their interpretation is to be calculated *mutatis mutandis*, by a possible recombination of signs and associated uses seen earlier. This corresponds to the emphasis put by the compositionality principle on the role of syntactic combination.

However, the reformulation I am about to propose also differs from the compositionality principle in important respects. Both similarity and recurrence are *gradual* concepts, because one expression or meaning can be similar to another not only in various aspects, but also to varying degrees, and the recurrence of a pattern (the frequency of its previous occurrences or the strength of the memories we have of them) also comes in degrees. The linguistics of the mid-20th century was reluctant to use models involving gradualness, but since many human systems, especially mental ones, clearly show gradualness effects, it will be hard to avoid appropriate stochastic models in the long run. So my generalized principle can be formulated along the following lines:

(4) *The principle of generalized compositionality*

To achieve maximal understanding, if we want to express an idea *I*, then we had better use an expression *E* that is maximally similar to the most frequent expressions *E'* which, to our knowledge, express ideas *I'* maximally similar to *I*. Also, this is the strategy that a hearer assumes other speakers to use.

The reason why I dare call this principle a generalization of compositionality is that it also expresses that different forms that show some parallelism are interpreted in parallel ways. Under traditional compositionality, formal parallelism is restricted to constituent structure, and semantic parallelism is restricted to the identity of the functions that combine the constituents' meanings. Thus my reformulation of the compositionality principle, in addition to the gradualness inherent in it, also crucially lacks a reference to syntactic constituency. For example, in the example I have examined earlier, it is crucial that *Boston office* is similar to expressions frequently heard earlier, in the sense that, in those expressions, the first word was similar to *Boston* (not in the phonological sense, but in terms of being names of places) and the second was similar to *office* (again, not in terms of their form, but their function), and those expressions were interpreted analogously to *Boston office*. These similarities involve "constituents" (it does matter that *Boston office* is built up of two sub-expressions in a particular linear order), but the similarity between "constituent structures" is just one particular case of many types of similarity. For example, similarities between word forms (i.e., the morphological built-up of words) is often not restricted to linear, concatenative similarity. As a matter of fact, as we have shown elsewhere (Kálmán & al. 2005), it would be wrong to consider linear and concatenative morphology as "normal" and qualify all other cases as deviant.

### 2.3 *Analogy-based semantics*

My claim is that the principle of generalized compositionality in (4) not only amends the weaknesses inherent in the traditional concept of compositionality, but also solves problems like those presented in 1. That is, it offers an alternative, more appealing approach to how linguistic and background knowledge contribute to possible interpretations, and to how ambiguities arise in various contexts.

To explore this alternative, I will assume that, as long as we lack evidence to the contrary, there is a single stock of knowledge for "linguistic" and "background" knowledge, i.e., an ontology that comprises knowledge about both linguistic and extra-linguistic entities. Syntactic structures are based on surface utterances, no "underlying" or "abstract" structure is posited. No "translation" occurs at all; the structure of an utterance plays a role only inasmuch as it expresses the formal *similarity* of one utterance to the other.

Instead of translating expressions into a logical language, we just retrieve information associated with the linguistic entities perceived, be it linguistic or extra-linguistic in character. Thus, information, linguistic and non-linguistic alike, is activated through *association*. In the examples quoted in 1.1, the relevant information comprises everything related to *Boston* and *office* and their syntactic arrangement (i.e., this particular type of compound). Those pieces of information get activated, together with all the information originating from the rest of the utterance and the utterance context. In the case of the problem in 1.2, the relevant information that gets invoked includes the formal and functional aspects of other instances of clauses embedded in (negative) predicates, other instances of NP disjunctions, etc.

#### 2.4 *The abduction process*

Independently of whether we assume two separate modules like (Hobbs & al. 1993) or a unique network-like knowledge base, the abduction process must be an optimization of continuous variables, since plausibility is gradual. The task is a quite complex one: as I have pointed out earlier, we have to retrieve a *maximal consistent set of maximally plausible* abduced premises for a set of conclusions. For (Hobbs & al. 1993), the conclusions are constituted by the “translation” of the input utterance; in my view, they must consist of all we might consider *empirically maximally certain*, i.e., what has been perceived, in whatever way we model that. For the example of *Boston office*, this means that we can only take it for certain that the words *Boston* and *office* have been uttered in this order (plus whatever contextual information we want to take into account).

To be sure, plausibility must rest on probability or frequency: the more probable or frequent a state of affairs is, the more plausible it is to assume it (at least with a certain probability) given that we have some evidence for the truth of something that follows from it (again, the probability that this evidence is reliable may vary). Our aim is to maximize the joint probability of the entire set of propositions, perceived and abduced, taken together. We also have to take into account the *synergy* of perceived facts and plausible assumptions, in the sense that taking the joint probability of several antecedents taken together may yield a different result from just looking at their probabilities one by one. For example, take our example *The Boston office called*. Hobbs & al. (1993) argue that the most plausible assumption to make is ‘someone working at the Boston office placed the call’, because only humans can make telephone calls, and relating a human to an office can be done most plausibly by assuming that the person works there (how exactly this is achieved by (Hobbs & al. 1993) is irrelevant here). However, *call* is itself polysemous (animals or even machines can call as well). On the other hand, even if calling was uniquely human, this would not ensure for the above abduction to be correct: *The Boston office is on holiday* would

not be interpreted (if it is interpretable at all) as ‘someone working at the Boston office is on holiday’ but rather as ‘everyone working at the Boston office is on holiday’. It is only the synergy of all more or less plausible premises, i.e., their high joint probability, that makes the abduction suggested in (Hobbs & *al.* 1993) really plausible.<sup>2</sup>

These conditions suggest a solution in the spirit of *Bayesian networks*—e.g., (Pearl 1985) or *Markov logic networks* (Richardson & *al.* 2006). Calculations using both of these models are extremely complex, which suggests that *massively parallel* computational mechanisms are required in order to deal with them with reasonable resources. The model that I will propose shortly, in 2.6, is intended to solve this.

### 2.5 Accounting for NP disjunction

In accordance with the alternative approach proposed here, when interpreting sentences like (1) in 1.2, we do not depart from their genuine (surface) syntactic structure, and try to derive their interpretations from what the sentence actually looks like. In particular, we must take into account how sentences with a negative content and, in particular, the arguments of negative predicates are interpreted in other cases. On the other hand, we also have to look at how the disjunction and, in general, the co-ordination of NP’s is interpreted in other cases. Interpretation has to proceed by analogy to all those other cases.

This approach has two obvious consequences which I think are desirable. First, it predicts that uncertainty may arise in the interpretation of the utterances in question if there are analogous structures with conflicting interpretations. Such uncertainties are hard to explain if one assumes a compositional semantics producing a logical form, given the fact that, informally speaking, we are dealing with the simplest syntactic structures and the simplest logical connectives. Second, it predicts that the interpretation of such utterances may show peculiar differences from one language to the other, even if independent arguments for structural differences would be hard to find.

In sum, I will assume that all the examples discussed in 1.2 and here contain disjunctions of noun phrases (i.e., we must not convert the disjunctions into sentential ones), embedded into (explicit or implicit) negation (i.e., there is no “scope ambiguity” involved). The idea is somewhat similar to Scha’s (1981), who argues against reducing plural or co-ordinated arguments to quantificational structures, and for conceiving of them as marking *predication about collections*. That is, when using a disjunction of noun phrases, one talks about collections the

<sup>2</sup>There are also additional propositions to be abduced that (Hobbs & *al.* 1993) fails to mention. For example, the person placing the call is not just anyone working there, but one who is somehow entitled to represent the Boston office; the office must belong to a company uniquely identifiable as familiar to both the speaker and the audience; and so on.

members of which are “alternatives” of each other (in some sense of the word), while the conjunction of noun phrases refers to collections that the predicate applies to “jointly” (in some sense of the word).

It is well-known that the exact interpretation of predicating about collections is, to a large extent, underdetermined by the linguistic structure used. One consequence of this is that ample space is left for contextual influences and uncertainty. In the case of “conjunctive” collections, possible interpretations include collective, distributive and cumulative readings (cf. Scha 1981); the possible interpretations of “disjunctive” ones are different: they express “choice” in a largely underspecified sense (including free choice, uncertainty and variation).

Obviously, the preferred interpretation of such an underspecified interpretation instruction may vary depending on what the disjunctive coordination is embedded into. For example, consider:

- (5) *I closed the door or the window.*
- a. 'I (always) closed some opening, I forgot/it does not matter if it was the door or the window'
  - b. 'Sometimes I closed the door, sometimes the window'
- (6) *Close the door or the window.*
- a. 'Close that opening, whether it is the door or the window'
  - b. 'Close either the door or the window'
  - c. 'Always close an opening, sometimes the door, sometimes the window'

The same effect underlies, I believe, the well-known puzzle, originally noted in (Kamp 1973), which involves disjunction embedded in a permission predicate:

- (7) a. *You can have soup or cookies.*  
       '(7a)'  $\models$  'you can have soup'  
       '(7a)'  $\models$  'you can have cookies'
- b. *You can have soup.*  
       '(7b)'  $\not\models$  'you can have soup or cookies'

Although Fox (2007) and others have argued that the entailments in (7a) are in fact conversational implicatures, they are not cancellable as conversational implicatures are supposed to be, which makes this type of analyses untenable. Instead, we must explain them by looking at similar structures, and reconstruct what sense people can make out of a disjunctive collection in a permission context. It is easy to see that, in this special type of contexts, predication about such disjunctive collections must be interpreted in nearly the same way as if they were conjunctive (although exclusiveness may be understood):

- (7') *You can have soup or cookies.*  
       'You can have whichever you want (but not both)'



As a matter of course, this sentence can also be interpreted by taking the disjunction to indicate uncertainty ('You can have food, but I don't know which kind is available'), it is simply less probable or plausible for real-world reasons.

As opposed to the permission context, the negative context easily yields both types of readings. The reading in (1'a), with "wide-scope negation", corresponds to the free-choice reading, whereas the reading in (1) corresponds to uncertainty:

(1) *He didn't close the door or the window.*

- a. 'He didn't close them; whichever you pick, it is true that he didn't close it'
- b. 'He didn't close one of them; I don't know which (or maybe both)'

(Szabolcsi 2002) claims that the Hungarian counterpart of (1) is less ambiguous than the English sentence. In particular, while reading (a) is preferred in English (if there is a preferred reading), reading (b) seems preferred in Hungarian. My own informal survey shows that, in fact, both readings exist in Hungarian, too, and they are more or less prominent depending on the context. (Note the difficulty of capturing how contextual factors can help disambiguation if the different readings are due to different syntactic structures.) For example, consider:

(8) *Nem csuktad be az ajtót vagy az ablakot.*

not closed-you in the door-acc or the window-acc

'You didn't close the door or the window'

- a. When explaining the source of the draught:  
'you left both the door and the window open'
- b. When explaining the source of the cold temperature:  
'you left the door or the window (or both) open'

In (a), the preferred interpretation is like in (1'a), because draught is due to two things open simultaneously, whereas (b) is more likely to be interpreted as (1'b), because one opening is sufficient for explaining the cold temperature.

If the difference between English and Hungarian is not as prominent as (Szabolcsi 2002) claims, then the syntactically grounded explanation faces a serious problem, because slight biases or weak tendencies are hard to account for in a theory of syntax like transformational generative grammar. Obviously, since we all believe that natural languages are systematic, it is essential that we look for parallel phenomena that can be related to the interpretation of disjunctions in negative contexts.

One obvious candidate for such a parallel phenomenon is the behaviour of *indefinites* in negative contexts since, in the logical sense, indefinites can be considered abbreviations of potentially infinite disjunctions. Indefinites can be interpreted in pretty much the same way as disjunctive collections:

- (9) a. *He closed a window.*  
 b. *You can have a cookie.*  
 c. *I bought a car.*

The interpretation of “choice” and/or “uncertainty” varies from one sentence to the other: a genuine free-choice reading is only available in (9b); in the case of the indefinite in (9a), the “choice” and the “uncertainty” interpretations are almost undistinguishable. The sentence in (9c) seems to have an “uncertainty” reading only, namely, the addressee reading, there is no “choice” whatsoever involved, only uncertainty, namely, the addressee cannot be certain what car the speaker bought.

When an indefinite occurs in a negative context, it usually has both a “choice” and an “uncertainty” interpretation:

- (10) *I didn't buy a car.*  
 a. 'I bought no car' (choice)  
 b. 'There is some car (it does not matter/you may not know which) I didn't buy' (uncertainty)

But English and Hungarian sharply differ in this respect:

- (11) *I didn't buy a car.*  
 (12) a. # *Nem vettem egy autót.*  
           not bought-I a car-acc  
 b. <sup>OK</sup> *Nem vettem autót.*  
           not bought-I car-acc  
           'I didn't by a car (= I bought no car)'  
 c. <sup>OK</sup> *Nem vettem meg egy autót.*  
           not bought-I pref a car-acc  
           'I didn't by one of the cars, there is a car I didn't buy'

That is, in Hungarian the “choice” interpretation of indefinites within a negative context can only be achieved by using *bare nominals* like *autót* 'car-acc' in (12b); a genuine indefinite only tolerates the negative context in an “uncertainty” reading, as in (12c), and only if a prefixed version of the verb is available (in this case, *meg + vettem* 'pref-bought-I' is the prefixed version of *vettem* 'bought-I', and the prefix is relegated to a post-verbal position because of the presence of negation).

Under an analysis like Szabolcsi's (2002), indefinites and disjunctions share the property that they avoid “having smaller scope than negation” (except maybe certain circumstances). But such an analysis would predict, e.g., that (8a) should be as bad as (12a), which is far from true. On the other hand, if one interprets the relevant sentences based on analogies with other structures, then

the argument goes as follows. In English, free-choice indefinites in negative contexts are highly frequent (cf. (10a)), whereas in Hungarian they are much rarer (given the unacceptability of sentences like (12a)). This predicts a stronger tendency for a similar distribution of “choice” vs. “uncertainty” readings in general in the respective languages, including the interpretation of predication about disjunctive collections. The effect on the readings of disjunctions can well be a slight one, which cannot be obtained in a theory that attributes the difference to features that trigger or block various transformations. As a matter of fact, one could even give an estimation of the degree of these tendencies if one had a sufficiently rich semantically annotated corpus.

## 2.6 *Machinery: An associative memory model*

As I have suggested in section 2.4, that solution to the logical problem outlined in 1.1 requires a formal model of *association* capable of embodying parallel stochastic algorithms. Such networks are well-known in the history of artificial intelligence, from the semantic nets of (Schank & al. 1969; Schank 1975, 1982) through Bayesian networks (e.g., Pearl 1985) to the countless versions of neural networks (for an overview see, e.g., Arbib 1995). It is not clear, however, how the type of analogical reasoning that I have outlined in the previous sections can be implemented in such systems. Purely symbolic networks such as Schank’s can model associations between concepts, but they do not make it possible to decompose and create new combinations of concepts by analogy of combinations seen earlier; purely neural networks, on the other hand, which do not contain explicit representations of concepts, are ill-understood, and it is not clear whether and how they can solve a particular logical problem like the one I am concentrating on.

For this reason, I am proposing a *hybrid* conceptual/neural network, which stores *aspects* or *properties* of concepts/experiences/memories, together with their *probability* (depending on how frequently they have been activated, i.e., observed or used). Novel combinations (new “concepts”, so to say) can arise in such a network, because we can interpret the simultaneous activation of properties as their recombinations. (For the problems discussed in this paper, the relevant properties include uses of place names and *office*, properties of compound nominals in English, negative predication, predicating about collections, “disjunctive” and “conjunctive” collections, and so on.) Since we intend to model association, the algorithm that operates on the network is *activation spreading*.

Obviously, the system is capable of *learning* by incorporating new properties (never observed earlier) and by updating the frequency of a property whenever it gets activated (for any reason, either by direct observation or by internal use). The main difference between this system and predecessors is that the *connections*, i.e., associations between properties, do not ever change: neither

new connections nor “connection weights” can be introduced (except when a new property gets incorporated in the network). Associations only exist between more and less *general* properties, just like in a classical *generalization network* (e.g., Hispanicus 1947; Levinson 1996), plus *inhibitory links* exist between more or less incompatible properties. The role of probability (frequency) is that the spreading of activation (and inhibition) is stronger from more frequent nodes. Crucially, activation also spreads more efficiently (with less loss) from more specific to more general nodes than the other way round. This corresponds to the idea that the decomposition of a more concrete experience (i.e., its association with its properties) is quasi-automatic, whereas the association from properties to the more concrete experiences that exhibit that property is less fluent.

The memory model outlined here is explained in detail in (Kálmán 2010). It has been successfully tested on relatively simple examples in morphology and syntax, but not yet in the much more complex realm of semantics. Therefore, for the moment, I can only speculate on how analogy can be made to work in the domain of semantics, as I have explained in sections 2.4 and 2.5.

### 3 SUMMARY

We have witnessed various problems with the dominant view of understanding. This view posits a level of “literal” or “linguistic meaning”, which is then supplemented by extra assumptions made by the audience using their “background knowledge”. But the borderline between “linguistic” and “background” knowledge is questionable. On the other hand, the principle of compositionality, which underlies modern semantics in general, and formal semantic theory in particular, also posits a level of “linguistic”, “context-independent meanings”, tacitly relegating all other information used in the production and understanding of linguistic signs to other modules (namely, “pragmatics”). So I had to reformulate this principle in order to bring it into harmony with the homogeneous view of understanding that I advocate.

The reformulation of the compositionality principle (which I have dubbed *generalized compositionality*), is based on a general principle of human linguistic communication, namely, *systematicity*, which means that similar forms tend to be associated with similar functions, and vice versa. (It should be noted that, as a consequence of this general principle, linguistic systematicity also means that different forms tend to be associated with different functions, and vice versa.) Since, as a consequence, the principle of generalized compositionality relies only on similarity and recurrence, it incorporates a *holistic* approach to linguistic signs: as opposed to the traditional, *atomistic* view, which aims at deriving all properties of complex signs from the properties of their constituent parts, the holistic view emphasises global similarities which, in principle, need not always be defined

recursively, as a similarity of “constituent structure” and the similarity of the corresponding “constituents”.<sup>1</sup>

## REFERENCES

- Arbib, M. A. (ed.), 1995, *The Handbook of Brain Theory and Neural Networks*, Cambridge (MA), MIT Press.
- Fox, D., 2007, Free choice and the theory of scalar implicatures. In U. Sauerland and P. Stateva (eds.), *Presupposition and implicature in compositional semantics*. Houndmills, Basingstoke, Palgrave Macmillan, 71–120.
- Gendler Szabó, Z., 2000, *Problems of compositionality*. New York, Garland.
- Grice, P. 1967, William James lectures on logic and conversation. In D. Davidson and G. Harman (eds.), *The Logic of Grammar*, Encino, Dickenson, 64–75.
- Petrus Hispanicus, 1947, *Summulae logicales*. Rome. (Edited by I. M. Bocheski; original from c. 1239)
- Hobbs, J. R., Stickel, M. E., Appelt, D. E. and Martin, P. A., 1993, Interpretation as abduction. *Artificial Intelligence* 63 (1–2), 69–142.
- Kálmán, L., forthcoming, Analogical reasoning using an associative memory model. Manuscript.
- Kálmán, L., Rebrus, P. and Törkenczy, M., 2005, Hungarian linking vowels: An analogy-based approach. Unpublished manuscript. 2nd Old World Conference in Phonology (OCP2), University of Tromsø, Center for Advanced Study in Theoretical Linguistics (CASTL), Tromsø, Norway.
- Kamp, H., 1973, Free choice permission. *Proceedings of the Aristotelian Society* 74, 57–74.
- Levinson, R. A., 1996, General game-playing and reinforcement learning. *Computational Intelligence* 12, 155–176.
- Pearl, J., 1985, *Bayesian networks: A model of self-activated memory for evidential reasoning*. Technical Report CSD-850017, Los Angeles, UCLA.
- Richardson, M. and Domingos, P., 2006, Markov logic networks. *Machine Learning*, 62, 107–136.
- Scha, R., 1981, Distributive, Collective and Cumulative Quantification. In *Formal Methods in the Study of Language*, Amsterdam, Mathematisch Centrum, 483–512.
- Schank, R. C. (ed.), 1975, *Conceptual information processing*. Amsterdam, North-Holland.
- Schank, R. C., 1982, *Dynamic Memory*. New York, Cambridge University Press.
- Schank, R. C. and Tesler, L. G., 1969, A conceptual parser for natural language. *Proc. IJCAI-69*, 569–578.
- Szabolcsi, A., 2002, Hungarian disjunctions and positive polarity. In I. Kenesei (ed.), *Approaches to Hungarian* 8. Szeged, SzTE.
- Szabolcsi, A. and Haddican, B., 2004, Conjunction meets negation: A study in cross-linguistic variation. *Journal of Semantics* 21(3), 219–249.
- Zadrozny, W., 1994, From compositional to systematic semantics. *Linguistics and Philosophy* 17, 329–342.

<sup>1</sup>I am grateful to Zsófia Zvolenszky for her comments on the manuscript.

## The Treatment of Ordinary Quantification in English Proper

**Abstract.** In this paper we bring together some well-known lines of criticism directed at Montague Grammar, such as (i) taking a stilted, highly regulated variety of language as the object of inquiry; (ii) ignoring the meaning of content words; and (iii) the failure to treat hyperintensionals; and suggest a coherent, and we believe much simpler alternative based on structured meanings.

### INTRODUCTION

Among the many inventions that made P1Q such a monument of intellectual achievement was the insistence on presenting a significant fragment of *ordinary* English, as opposed to the semi-formalized (sometimes fully formalized) and regimented English-like sublanguages used in most works of philosophical logic at the time. Yet upon rereading the founding papers of Montague Grammar (MG), in particular (Montague 1970a, 1970b, 1973), all reprinted in (Thomason 1974), linguists brought up in a more descriptive tradition are inevitably struck by the stilted “English” of the fragments. The problem is not so much that the pioneering examples from *Every man loves a woman such that she loves him* to *John seeks a unicorn and Mary seeks it* could hardly be regarded as examples of ordinary language but that there has been an alarming lack of progress in this regard. Nearly forty years have passed, and papers discussing seminumerical puzzles like *At least three professors graded at most five exams from seven or fewer students* are abundant, while the interpretation of ordinary language makes little progress. It’s not that people couldn’t say things like *My grandmother has more plates than utensils* it’s just that they don’t.

In this paper, while still avoiding the rough and tumble of actual spoken English, we take a less regimented language variety, that of copy-edited journalistic prose, and investigate the semantics of *ordinary quantification* in this variety we

call *English proper* (rather than ‘proper English’ to avoid the association with primness and pedantry). In Section 1 we discuss how various uses of *for all* and *every* are to be sorted out, and describe informally how the dominant usage of *every* can be accounted for in a model that preserves most of the goals, but very little of the formal techniques, of MG. In Section 2 we suggest that the analysis of quantifiers and similar function words needs to be supplemented by a more detailed analysis of content words, if semantics is to cover more realistic examples—statements and inferences of the sort found both in ordinary English and in philosophical discourse. Since this brings into sharp focus the traditional puzzles concerning hyperintensionals, we consider these in some detail, and arrive at the conclusion that standard intensional theory is not tenable.

## 1 ORDINARY QUANTIFICATION

Our sample of universally quantified expressions is drawn from an American newspaper, the San Jose Mercury News (Merc). We manually analyzed 300 issues, totaling some 45 million words, to demonstrate that ordinary quantification is not at all close to the logical variety described in MG. As we shall see in Section 1.1, the data is dominated by a large number of other patterns. Since these are rarely amenable to a purely lexical, purely grammatical, or purely pragmatic treatment, in Section 1.2 we introduce a new Principle of Responsibility that forces us to look at such constructions more closely.

### 1.1 *Universal constructions in English*

In our corpus we find over 2400 occurrences of the strings *for all* and *For all*. Many of these could be called idiom chunks: *For all the glamour of aerial fish planting, it was a mass production money-maker [1]* clearly does not mean anything like  $\forall x \text{glamour}(x)(\dots)$ . A descriptive label such as *idiom*, *(partially) lexicalized expression*, or *snowclone* already hints at the necessity of stepping outside ‘pure’ semantics, bringing resources of a lexical or pragmatic sort on the issue at hand. In what follows, we will describe constructions as (partially) fixed patterns in the spirit of (Fillmore and Kay 1997), and ask what expressions like *[the Clarence Thomas hearings]*, *for all their import [1]* or *For all their efforts at parity and fairness, [NFL officials ...]* actually mean. This is not a problem somewhere on the fringes of the data—as a matter of fact, examples like these are considerably more frequent than those involving standard quantifier readings.

<sup>1</sup>For reasons of expository convenience we will often considerably simplify the raw examples, indicating inessential parts by [ ] wherever necessary. In doing so, we attempt to make sure the simplified example remains an instance of English proper, i.e. an example that could be produced by a reasonable writer of English and would be left standing by a reasonable copy-editor.

We define a *construction* as a string composed of nonterminals (variables ranging over some syntactic category) and terminals (fixed grammatical formatives and lexical entries) with a uniform compositional meaning, obtained by a fixed process whose inputs are the meanings of the nonterminals and whose output is the meaning of the construction as a whole. Completely productive and highly abstract grammatical patterns such as

- (1)  $NP\langle\alpha PERS\ \beta NUM\rangle VP\langle\alpha PERS\ \beta NUM\ \gamma TENSE\rangle$

and highly specified and almost entirely frozen idioms such as

- (2)  $NP\langle\alpha PERS\ \beta NUM\rangle kick\langle\alpha PERS\ \beta NUM\ \gamma TENSE\rangle the\ bucket$

will both be treated as constructions. On occasion, when we are interested in the substitution of one construction into another, it will be necessary to assign a grammatical category (defined as including morphosyntactic features specified in angled brackets) to the construction as a whole, so a syntactic theory roughly along the lines of GPSG (Gazdar *et al.* 1985) is presupposed. The combinatorial flexibility of the constructions suggests that more powerful theories, such as TAGs or LTAGs, (Joshi and Schabes 1997) may actually be a better choice, but for our current purposes, we are not particularly interested in transference across patterns, such as the phenomenon that the agreement portion of (2) is obviously inherited from that of (1). As a limiting case, entirely frozen expressions, i.e. those constructions that no longer contain open slots, like *go tell it to the Marines*, are simply taken as lexical entries, in this case, with meaning ‘nobody cares if you complain’. (The indexicals implicit in the imperative *go* and explicit in the paraphrase *you* do not constitute open slots in the sense we are interested in here.)

Viewed from this perspective, the standard case, found in many examples like *the law [ ] makes helmets mandatory for all motorcyclists and passengers* or *[ ] lowers the quality of life for all concerned* is yet another construction:

- (3)  $X\ for\ all\ N$

where *N* is some (bare) noun phrase and *X* is some predicative element, often a verb or VP, as in *lowers the quality of life*, but, perhaps surprisingly, more often a nominal or NP, as in *a model for all nations*.

In many cases, the adjacency of *for* and *all* appear accidental, as in *[ ] honored for all his work* or *sell [ ] for all that the market will bear* even though in some of these cases the standard analysis ( $\forall x his\_work(x) \dots$  or  $\forall x market\_will\_bear(x) \dots$ ) remains *prima facie* available.



Turning to *every*, of which there are about five times as many examples, here the dominant pattern is indeed one that lends itself to analysis in the standard terms: *every Californian with a car phone*, *every case*, *every famous star*, ... Remarkably, about 20% of these are time adverbials, *every day*, *every week*, *every time*, *every night* and so on. There are considerably fewer constructions with *every* than with *for all*, but they all share the property with (1) that they admit exceptions. In many cases, this is confirmed directly by the text: *every case, except that of Sen. Kennedy* []. In others, the text offers no overt exceptions, but it is clear that exceptions can be made: *every Californian with a car phone* except, of course, drivers of emergency vehicles...

Manual inspection of a large number of *every N<BAR I>* constructions makes clear that their meaning is really 'every non-exceptional N' rather than 'every N'—in fact, in the whole Merc corpus we could not find a single example of the latter. The results would have been very different with a corpus based on calculus textbooks, and we do not deny that the episodic reading routinely analyzed in MG exists, at least in a regimented variety of technical English—the claim is simply that it is not a part of English proper.

We should add here that we claim no originality in recognizing the problem, as the defeasibility of natural language statements has already given rise to a wide variety of *non-monotonic logic* approaches (for an overview see (Ginsberg 1987)) and the fact that generics admit exceptions is often viewed as one of their defining properties since (Jespersen 1924). If there is an original claim to be made in this area, it is that universal quantification, as the term is understood in predicate calculus, plays no role whatsoever in ordinary English or, indeed, in any natural language.

To put this finding in the harshest possible terms, PTQ fails to deliver on its major promise to treat quantification in ordinary English, concentrating on the jargon of mathematics instead. While subsequent work in the MG tradition such as (Moltmann 1995) and (Lappin 1996) have clearly recognized, and to some extent resolved the local problem of exceptionality, the global problem of dealing with the large variety of relevant constructions sampled here remains as acute as ever.

## 1.2 *The meaning of constructions*

Among the thousands of constructions used in English only a handful like *(go) tell it to the Marines* are amenable to a purely lexical treatment, and only a handful like *S → NP VP* are purely compositional. In between, there is a vast range of expressions containing one or more open slots, and our primary interest in doing semantics lies with interpreting these. There is a clear intuition that nonce phrases such as *California driver* differ from lexicalized forms such as *Rottweiler dog* only in that the latter is part of the lexicon, and the basic

components of its meaning is no longer surface accessible. California drivers are obviously humans with two properties that are true by definition, namely, that they live in California and that they drive a car, and many others that are derivable from these, such that they are above the California driver age limit or that they are featherless bipeds. In the formal model that we outline in Section 3, the meaning of nominals will be taken as a bundle (unordered conjunction) of predicates that correspond to the *essential* properties of the nominal in question. It is clear that both 'being Californian' and 'being a driver' are definitely part of the bundle of properties that form the base of the semantic model for California drivers, and we leave open the possibility that several other predicates are also part of the bundle. For example, it is hard to know whether the stereotype that California drivers are polite is part of the lexicon (viewed as a purely grammatical construct) or belongs in some nebulous encyclopedia of world knowledge. But to the extent speakers of English can and do pursue inferences on this basis, we view it as part of the task of semantics to account for these. We state this as our *Principle of Responsibility*:

The semantics of any expression must be fully accounted for by the lexicon and the grammar taken together.

The Principle of Responsibility is only slightly stronger than the standard Principle of Compositionality which takes the semantics of any expression to be determined by the semantics of its lexical components and by the grammatical way those are combined. The additional requirement it imposes is that the 'pragmatic wastebasket' remain empty at all times: it doesn't matter whether we call ordinary inferences grammatical, lexical, or pragmatic (and perhaps extragrammatical), the overall system needs to account for these, either in one specific component, or by means of tracing the inference process through several components.

Let us begin with the *for all NP<+DEF>, S* construction. Clearly, this means something like '*S*, in spite of the usual implications of *NP<+DEF>*'. In the case of *the glamour of aerial fish planting*, the implication that needs to be defeated is that glamorous things are restricted to the few, a notion incompatible with *mass production*. The lesson from the example is already clear: to make sense of the construction we need to use a great deal of lexical information. Without doing so, the clear difference between the acceptability of the Merc examples and *???For all their protein content, eggs are shaped so as to ease passage through the duct* would remain completely mysterious.

As for non-exceptionality, mathematics offers two significantly different formal reconstructions of this notion. One approach, exemplified in (4), relates non-exceptionality to probability:

(4) *The typical number is irrational*

and can be rephrased as ‘the set of exceptions has measure zero’. The other, exemplified in (5), relates to the satisfaction of no extra predicates:

(5) *The typical square matrix over  $Q$  has unequal eigenvalues*

Statements like (4) occur so frequently in mathematical discourse that they have a terminus technicus of their own: we say *almost all* [numbers are irrational]. Statements like (5) are interpreted more in terms of dimension than in terms of measure, and we speak of *lower dimension* when reducing these to more primitive notions.

These two approaches to non-exceptionality are not incompatible, but it may take very significant work to establish, as Martin-Löf (1966) did, that the statement *The typical binary string is not compressible* yields the same definition from both the measure-theoretic and the no-extra-predicates standpoint. Here we choose the second approach in light of the fact that there is no obvious way to define measure spaces over semantic objects like *legal cases* or *California drivers with car phones*.<sup>2</sup> Thus, to say that *Geraldo Rivera* [ [ ] reveals that he is an extremely attractive virile hunk of man who ] has had sex with [ ] every famous star in the entertainment industry [ ] is to say that for all  $x$  such that  $x$  has no extra properties beyond being a famous star in the entertainment industry, Geraldo Rivera has had sex with  $x$ .

A less clumsy translation, very much in the spirit of generalized quantification, would be to say that the property of having had sex with Geraldo Rivera is implied by the property of being a famous star in the entertainment industry, and this is what we adopt for ordinary quantification: we say that *every  $N$*  is the set of *typical* properties that  $N$  has, where typicality is defined in the lexical entry of  $N$ . Since having four legs is typical of donkeys, *every donkey has four legs* will be true by definition, and cannot be falsified by the odd lame donkey with three or fewer legs.

But if having four legs is an analytic truth for donkeys, what about counterfactuals where five-legged donkeys can appear easily, or the rather clear intuition, not disputed here, that being four-legged is a contingent fact about donkeys, one that can be changed e.g. by genetic manipulation? The answer offered here is that to reach these, we need to change the lexicon. Thus, to go from the historical meaning of Hungarian *kocsi* ‘coach, horse-driven carriage’ to its current meaning ‘(motor) car’ what is needed is the prevalence of the motor variety among ‘wheeled contrivances capable of carrying several people on roads’. A 17th century Hungarian would no doubt find the notion of a horseless coach just as puzzling as the notion of flying machines or same-sex marriages. The key

<sup>2</sup>Not only do we need a measure space, we would need substantive agreement that this particular measure is the one that is “natural” to the domain, just as Lebesgue-measure is agreed to be *the* natural measure for real numbers. While defining measures for semantic objects can be done many ways, arguing for any of these as being natural is much harder.

issue in readjusting the lexicon, it appears, is not counterfactuality as much as rarity: as long as cloning remains a rare medical technique we won't have to say 'a womb-borne human'.

A notable consequence of our definition by typical properties is that the translation of *every donkey* will not differ significantly from that of *any donkey*, *a donkey*, *donkeys* or even *the donkey*: the typicality restriction pertains to them all. This is as we want it for cross-linguistic purposes, since the clearly generic readings are not tied to the same varieties of quantified NPs in all languages.

To summarize what we have so far: *every man loves a woman* means neither  $\forall x \text{man}(x) \exists y \text{woman}(y) \text{loves}(x, y)$  nor  $\exists y \text{woman}(y) \forall x \text{man}(x) \text{loves}(x, y)$ —it means that woman-loving is a typical property of men, just as donkey-beating is a typical property of farmers. Importantly, it requires evidence beyond what is available in the example sentences to know whether farmers beat every donkey they can lay their hands on or just their own, and whether men love every woman or just one.

## 2 CONTENT WORDS

Legend has it that once in a semantics class a student asked Barbara Partee *What is the meaning of life?*, and she responded, after a moment of thought, by writing *life* on the blackboard. Since a key goal of the whole semantics enterprise is to provide a more satisfactory answer, we begin with analyzing two well known approaches. The first one, conventionally attributed to Koheleth, the author of Ecclesiastes, is that life is vanity, entirely devoid of meaning or purpose. According to Macbeth, "life ... is a tale told by an idiot, full of sound and fury, signifying nothing" (Act 5, Scene 5). Perhaps the most articulate exponent of this position is Schopenhauer, but we find many thinkers expressing the same idea before and after him.

The other well known answer is the religious one, that the meaning of life is to serve God. Somewhat surprisingly, given the magnitude and importance of the problem, this answer is relegated to a subordinate clause of a longer story concerned with something else, hidden in a book generally regarded minor, Isaiah 43.6-7, wherein the standard Judeo-Christian approach is spelled out as follows: "bring my sons ... every one that is called by my name: *for I have created him for my glory*, I have formed him; yea, I have made him." Lest the reader feel disappointed we should emphasize at the outset that our primary interest here is not so much with exegesis as with lexicography. Instead of attacking the major problem posed by the student and many before him, we merely seek a technical approach that at least makes it possible to formulate the traditional answers sketched above.

We see in Partee's witty response a deeper truth, namely that Montague semantics lacks entirely the resources to approach issues of word meaning. The

problem, from our standpoint, is not so much that we don't know the meaning of life; rather, the problem is that even if we did, we couldn't express it within the standard framework. Assume, for a moment, that the first answer is correct, that life has no meaning. Does this mean  $\text{life}' = \emptyset$ , and if so, how can this be derived along the lines proposed by Koheleth, from the observation that "All go unto one place; all are of the dust, and all turn to dust again"? Perhaps we need a subsidiary axiom that meaning is a permanent, unchanging and unchangeable thing, so that if something is not eternal it must be meaningless. But is it the object that must be eternal or is it its meaning, and how would we distinguish the two cases? Or assume, for the sake of argument, that the second answer is the correct one, that God has created Man for His glory. Really, what does this mean? Does it mean, on account of the masculine pronoun being used, that Woman is excluded? And what is *glory* so that even God cares to have more of it? The whole MG framework, which treats the meaning of everything other than a few function words as an unanalyzed set, is incapable of formulating, let alone resolving, such questions.

The issue is really not the meaning of the word *meaning*. The main question could be recast in many other ways, as an inquiry concerning the *purpose* of life, the *goal* of life, and so forth. The technical reconstruction of 'meaning' used by MG as the set of instances in this world or other possible worlds, is quite satisfactory. But what is the set of purposes, the set of goals, or even the set of living things? We do not wish to create a mystery where there isn't one, for there are perfectly reasonable commonsensical answers which accord well both with everyday and with philosophical usage, and every dictionary will have some version of these, stated in terms of a rather simple theory of lexical semantics wherein the meaning of nominals is conceived of as a bundle of properties.

Historically such a theory can be traced back to Leibnitz, the Schoolmen, and eventually to Aristotle. In contemporary semantics this idea is at the foundations of both the Semantic Web (there called Web Ontology Language or OWL) and of the influential WordNet approach to the lexicon (Miller 1983). Yet in contemporary philosophy such theories have little credibility since Russell's (1905) critique of Meinong, with (Parsons 1974) definitely remaining a minority view. The technical difficulty is in formulating inferences based on such dictionary definitions, because the theory that can sustain definitions for nouns such as *fox* composed of properties such as *animal*, *four-legged*, *red*, *clever* can also sustain definitions of inconsistent and/or non-existent objects, to which we now turn.

## 2.1 *Hyperintensionals*

Russell had two major objections: first, a technical one concerning the existence predicate. Let us take any nonexistent object of Meinong's Jungle.

such as the gold mountain. Such an object doesn't exist, but if Meinong is right, and any combination of properties can be construed as an object, the existent gold mountain is also part of the Jungle, and as such it is not only gold, and mountain, but also existent, contradicting our earlier assertion that it was nonexistent. This objection is easily defused by splitting the existence predicate in two, 'exists in the mind', and 'exists in reality', and exempting the second meaning from the class of predicates that can be used to describe (or in the mind, create) noun objects. Russell's second point, concerning inconsistent bundles of properties, such as triangular circles, was that they violate the law of non-contradiction. Before going any further, let us consider some examples.

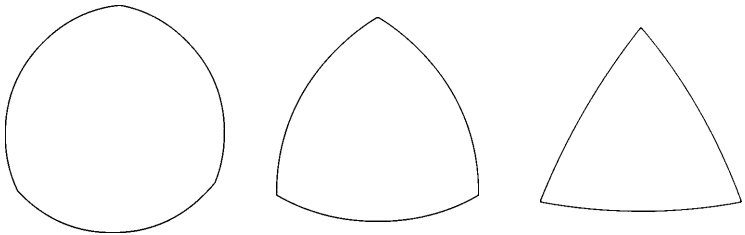


Figure 1. The Reuleaux triangle and its cousins

The figure shows on the left a slightly triangular circle, and on the right a slightly circular triangle. Whether the object in the middle, known as the “Reuleaux triangle”, is considered a triangle, a point of view justified by its having three distinct vertices, or a circle, a point of view justified by its having a constant diameter, is a matter of perception.

What is clear from the linguistic standpoint is that adjectives like *slightly*, *seemingly*, *very* attach to adjectives like *circular*, *triangular*, *equal* that have a strict mathematical definition just as easily as they attach to adjectives like *red*, *large*, *awful* that lack such a definition. Clearly, what these adjectives modify is the “everyday” sense of these terms—the mathematical sense is fixed once and for all and not subject to modification. Just as we were interested in the everyday sense of *all* and *every* and found that these are distinct from the standard mathematical sense taken for granted in MG, here we are interested in the ordinary sense of *circular*. Working backwards from typical expressions like *circular letter*, *circular argument*, we find that the central aspect of the meaning is not ‘a fixed distance away from a center’ or even ‘fixed diameter’, but rather ‘returning to its starting point’, ‘being cyclic’.

In these examples, the morphologically primitive forms are nominal: the adjectival forms *circular*, *triangular* are clearly derived from *circle*, *triangle* and not the other way around. Since derivations of this sort change only the syntactic category of the expression but preserve its meaning, we can safely conclude that

*circle* in the everyday sense is defined by some finite conjunction of essential properties that includes ‘being cyclic’ and that the mathematical definition extends this conjunction by ‘staying in an (ideal) plane, keeping some (exact) fixed distance from a point’. Similarly, *triangle* simply means ‘having three angular corners’ rather than the exact configuration of points and lines assumed in geometry. Taking these notions together, the predicate bundle of *triangular circle* will contain the properties *curve*, *cyclic*, *has(three vertices)* and all three shapes depicted above will fit this definition.

A more general example of the same mechanism is provided by the adjectival modification of proper names. Who is *the Polish Shakespeare* that the “*Looking for the Polish Shakespeare*” *Contest for Young Playwrights* wants to find? Clearly, not some British subject born in Stratford-upon-Avon but a brilliant playwright who is a Polish national. Once we recognize that adjectival modification is not simply a conjunction of some new property to the set of essential properties, but one that may interpose a higher predicate that is only implicit (as is *nationality* in this case), a whole range of otherwise puzzling constructions become transparent. Altogether, English proper has far fewer hyperintensional constructions than hitherto assumed: certainly triangular circles and immaculate conceptions do not give rise to logical contradictions. This renders the well-known problem with hyperintensionals in the MG account of opacity far less urgent, as we now only have to deal with cases in which the essential meaning of the adjective is in strict contradiction to the essential meaning of the noun it modifies, and the latter is given by a single conjunct. Thus we need to consider examples like

(6) *Mondays that fall on Tuesdays*

(7) *Mondays that fall on Wednesdays*

Does it follow that a (rational) agent who believes in (6) must also believe in (7)? While the matter is obviously somewhat speculative, we believe the answer to be negative: if we learn that *Peter believes Mondays can be really weird—he actually woke up to one that fell on Tuesday* it does not follow that he also believes himself to have woken up on one that fell on Wednesday. Since he has some sort of weird experience that justifies for him a belief in (6), he is entitled to this belief without having to commit himself to (7), as the latter is not supported by any experience he has. If this is so, the hyperintensional problem is still relevant, and the intensional treatment of opacity cannot be maintained for all cases. But if it cannot be maintained for all cases, there does not seem to be a compelling reason to maintain it at all, since a simpler alternative treatment, based on structured meanings (Cresswell 1985), is available (see (Ojeda 2006) for detailed argumentation why a non-intensional treatment is to be preferred both on grounds of simplicity and grounds of adequacy). Being algebraic, the

structured meanings we use (see (Kornai 2009)) are somewhat different from those used by Cresswell, but all the reasoning that leads to structured meanings remains applicable. In particular, the theory presented here puts fiction on a par with factual discourse: when we assert truthfulness we do this relative to a particular model, so that *Anna Karenina commits suicide* can be true while *Jan Valjean commits suicide* is false, relative to their respective models.

According to Thomason (1977) a heavy price must be paid for adopting a semantic theory based on structured meanings: in brief, such a theory is incompatible with a nontrivial theory of truth. In his discussion of Montague (1963), Thomason argues that any direct theory of propositional attitudes is bound to be caught up in Tarski's (1935) Theorem of Undefinability. However, as Thomason is careful to note, the conclusion rests on our ability to pass from natural language to the kinds of formal systems that Tarski and Montague consider: first order theories with identity, strong enough to model arithmetic. Tarski himself was not sanguine about this: he held that in natural language "it seems to be impossible to define the notion of truth or even to use this notion in a consistent manner and in agreement with the laws of logic". Russell held similar views, calling natural language "a rough and ready instrument incapable of expressing Truth with a capital T".

To replicate Tarski's proof, we first need to supplement natural language with variables. The basic idea—to formalize the semantics of a predicate like *subject owns object* by a two-place relation  $\zeta(s, o)$ —is fairly standard (although there are significant alternatives that do not rely on variables at all). But the proposed paraphrases for first order formulas, such as replacing  $\forall x[\exists y\zeta(x, y) \rightarrow \exists z\zeta(z, x)]$  by *for everything x, either there is not something y such that x owns y or there is something z such that z belongs to x* clearly belong to an artificially regimented extension of English, rather than to English proper. Second, we must assume that the language can sustain a form of arithmetic, e.g. Robinson's Q.

The universality of natural language (or English proper) as a means of supporting logical or arithmetic calculi, is highly doubtful, and using Q we can pinpoint the source of these doubts more narrowly: several of the axioms in Q appear untenable for natural language. Interestingly, the key issues arise long before we consider exponentiation (a central feature for Gödel numbering) or ordering. Q comes with a signature that includes a successor  $s$ , addition  $+$ , and multiplication  $\cdot$ . By Q2 we can infer  $x = y$  from  $sx = sy$ , Q4 provides  $x + sy = s(x + y)$ , and Q6 gives  $x \cdot sy = xy + x$ . All of these axioms are gravely suspect in light of the following Principle of Non-Counting:

If  $\alpha p^n \beta \in L$  for  $n > 4$ ,  $\alpha p^{n+1} \beta \in L$  and has the same meaning

The notion that linguistic structures are non-counting goes back at least to Chomsky (1965, 55) and pervades every variety of syntax that we know



of.<sup>3</sup> There are many ways we can *start* counting in natural language: we can look at quotations of quotations (*Joe said that Bill said...*), emphasis of emphasized material (*very very...*), but there is not a single way that takes us very far—whichever way we go, we reach the top in no more than four steps, and there Q2 fails. Since on this key point, the semantics of natural language expressions parts with the semantics of mathematical expressions, the empirical underpinnings of the Tarski/Montague/Thomason argument are missing: there is no loss entailed by the use of structured meanings, as there was no chance to maintain a Tarski-type theory of truth in natural language to begin with.

### 3 CONCLUSIONS

Historically, MG started out as an ambitious but quite reasonable research program, explicitly moving away from the stilted examples of an earlier generation of logic textbooks toward ordinary natural language. Indeed, many of the pivotal examples motivating much subsequent work, e.g. Bach-Peters sentences, have an immediate impact, clearly comprehensible to any native speaker. We believe that increasingly arcane examples with increasingly contrived readings reasserted themselves as the primary focus of interest simply because MG and its modern descendants, starting with Dowty (1979), concentrated on elucidating the semantic analysis of those expressions for which the underlying logic had the resources. Since Montague's intensional logic *IL* includes a time parameter, in depth analysis of temporal markers (tense, aspect, time adverbials) becomes possible. But as long as the logic lacks analogous resources for space, kinship terms, sensory inputs, or obligations, this approach has no traction, and heaping all these issues on top of what was already a computationally intractable logic calculus has not proven fruitful. The goal of this paper was to drastically realign the focus of formal semantics from interesting puzzles and a Turing-complete higher order intensional apparatus to data of the simple and frequent kind that is likely to dominate the language acquisition process. We must crawl before we walk, and if we cannot account for the data that is likely seen during language acquisition our account of more complex phenomena is in doubt.

### ACKNOWLEDGMENTS

This paper builds extensively on ideas presented in Kornai (2008). We thank Károly Varasdi (HAS Institute of Linguistics), Anna Szabolcsi (NYU), and Zoltán Gendler Szabó (Yale) for penetrating comments on earlier drafts.

<sup>3</sup>Some limited counting, such as the building of binary (and perhaps ternary) feet, is generally assumed in phonology.

## REFERENCES

- Chomsky, N., 1965, *Aspects of the Theory of Syntax*. Cambridge (MA), MIT Press.
- Cresswell, M. J., 1985 *Structured Meanings*. Cambridge (MA), MIT Press.
- Dowty, D., 1979, *Word Meaning and Montague Grammar*. Dordrecht, Reidel.
- Fillmore, Ch. and P. Kay, 1997, *Berkeley Construction Grammar*.  
<http://www.icsi.berkeley.edu/~kay/bcg/ConGram.html>
- Fine, K., 1985, *Reasoning with Arbitrary Objects*. Oxford, Blackwell.
- Gazdar, G., E. Klein, G. K. Pullum and I. A. Sag, 1985, *Generalized Phrase Structure Grammar*. Oxford, Blackwell.
- Ginsberg, M. L. (ed.), 1986a, *Readings in Non-monotonic Reasoning*. San Mateo(CA), Morgan Kauffman.
- Jespersen, O., 1924, *The Philosophy of Grammar*. London, George Allen and Unwin.
- Aravind K. Joshi, A. K. and Y. Schabes, 1997, Tree-adjointing grammars. In *Handbook of Formal Languages*, volume 3: Beyond Words. Berlin, Springer.
- Kornai, A., 2003, The algebra of lexicography. In G. Jäger and J. Michaelis (eds), *Proceedings of the 11th Mathematics of Language Workshop*, FoLLI Lecture Notes in Artificial Intelligence. Berlin, Springer Verlag.
- Kornai, A., 2008, *Mathematical Linguistics*. Berlin, Springer Verlag.
- Lappin, Sh., 1996, Generalized quantifiers, exception phrases, and logicity. *Journal of Semantics* 13, 197–220.
- Martin-Löf, P., 1966, The definition of random sequences. *Information and Control* 6, 602–619.
- Miller, G. A., 1995, Wordnet: a lexical database for english. *Communications of the ACM* 38, 39–41.
- Moltmann, F., 1995, Exception phrases and polyadic quantification. *Linguistics and Philosophy* 18, 223–280.
- Montague, R., 1963, Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability. *Acta Philosophica Fennica* 16, 153–167.
- Montague, R., 1970a, English as a formal language. In (Thomason 1974), 188–221.
- Montague, R., 1970b, Universal grammar. *Theoria* 36, 373–398.
- Montague, R., 1973, The proper treatment of quantification in ordinary English. In (Thomason 1974), 247–270.
- Ojeda, A., 2006, Discontinuous constituents. In *Elsevier Encyclopedia of Languages and Linguistics*. Amsterdam, Elsevier.
- Parsons, T., 1974, A prolegomenon to Meinongian semantics. *The Journal of Philosophy* 71, 561–580.
- Russell, B., 1905, On denoting. *Mind* 14, 441–478.
- Tarski, A., 1956, The concept of truth in formalized languages. In A. Tarski, editor, *Logic, Semantics, Metamathematics*. Oxford, Clarendon Press, 152–278.
- Thomason, R. H. (ed.) 1974, *Formal Philosophy: Selected papers of Richard Montague*. New Haven, Yale University Press.
- Thomason, R. H., 1977, Indirect discourse is not quotational. *The Monist* 60, 340–354.

# Atomic Descriptions in Dynamic Predicate Logic

**Abstract.** We introduce a version of dynamic predicate logic (*DPL*, see (Groenendijk & *al.* 1991) as a framework to model in a compositional way the dynamics of definite descriptions put forward in (Lewis 1979). The resulting system, *dynamic predicate logic with descriptions (DPLD)* borrows the concept of a referent system from an upgraded version of *DPL* introduced in (Groenendijk & *al.* 1996). It is an interesting feature of *DPLD* that not only formulas but also individual terms are capable of updating discourse information.

## 1 INTRODUCTION

David Lewis, in his seminal paper (Lewis 1979) that served as a philosophical basis for the project of dynamic semantics, put forward a cluster of linguistic phenomena that deserve dynamic treatment. One of them is the anaphoric use of definite descriptions. Strangely enough, in more than thirty years since Lewis's paper this phenomenon has apparently not been studied within the framework of dynamic semantics.<sup>1</sup>

According to Lewis, examples like “The dog got in a fight with another dog” show that “[i]t is not true that a definite description ‘the *F*’ denotes *x* if and only if *x* is the one and only *F* in existence. Neither is it true that ‘the *F*’ denotes *x* if and only if *x* is the one and only *F* in some contextually determined domain of discourse.” (348.) Instead, “[t]he proper treatment of descriptions must be more like this: ‘the *F*’ denotes *x* if and only if *x* is the most salient *F* in the domain of discourse, according to some contextually determined salience ranking.” (ibid.)

<sup>1</sup>On the other hand, it has been studied to a certain extent within the *DRT* framework; see (Eijck & *al.* 1997, 189ff.) The *DRT* treatment is completely different from ours.

This view contradicts the classical view on descriptions dominant in logic ever since Frege and Russell at crucial points.

Salience may be determined by the context of utterance or antecedent expressions. Both can be found in Lewis's following example:

- (1) The cat is in the carton. The cat will never meet our other cat, because our other cat lives in New Zealand. Our New Zealand cat lives with the Cresswells. And there he'll stay, because Miriam would be sad if the cat went away.

In Lewis's analysis, the referent of the first occurrence of "the cat" is determined by external conditions, and the second one refers back to the first. Then, the term "our other cat" introduces another discourse referent, and "our New Zealand cat" refers back to "our other cat", which has meanwhile been associated with the term "New Zealand".

All of these four description occurrences deserve special attention. The semantic phenomena involved in this sample discourse are very complex and have far-reaching implications, only a small portion of which can be discussed in this paper. In particular,

- (a) we are going to focus on atomic descriptions like "the cat" or "the man", since they seem to be more capable of linking anaphorically, while complex descriptions tend to behave in the Frege–Russell way;
- (b) we attempt to model the anaphoric use of atomic descriptions, like "the cat" referring back to "a cat" and "the one that lives with us" referring back to "the cat" in the discourse "We have a cat that lives with us, and another one that lives in new Zealand. The one that lives with us is old";
- (c) we also attempt to model the way atomic descriptions take discourse referents from external sources (or even introduce new discourse referents) when they have no salient antecedent, like "the cat" in the opening sentence of Lewis's example "the cat is in the carton":

The technical problem to be solved is that of representing the way salience information is gathered, updated and put to use in a discourse. In our approach, salience is a relation between discourse referents and predicates. Since the former are semantic entities, while the latter syntactic ones, this is a mixed type of information. Clearly, a predicate itself is a better choice as an object of salience information than its extension. But we cannot rule out the possibility that intension plays a role in salience ranking, as the following example may suggest:

- (2) A bachelor has a date with an old maid. ??The man is nervous. The woman is bored.

Whether and how anaphoric linking in this example works is beyond the scope of this paper.

We take Lewis's statement "[i]t is not true that a definite description 'the  $F$ ' denotes  $x$  if and only if  $x$  is the one and only  $F$  in existence" as an exaggeration. An account of definite descriptions that takes the uniqueness and existence conditions attached to "the smallest prime number" or "the author of *Counterfactuals*" as merely extreme cases of salience ranking would be highly counterintuitive. The use of descriptions that involves uniqueness and existence (henceforth, we refer to a description interpreted this way as a *classical description*) is just as legitimate as the anaphoric one, and it may even have priority over that. But descriptions, like so many other natural language expressions, are tools that—in spite of being of the same form—have different uses. In this paper we attempt to model one of these uses.

## 2 THE DYNAMICS OF ATOMIC DESCRIPTIONS

Cross-sentential anaphora is one of the two semantic challenges that *DPL* took on (the other being the anaphoric relations in donkey-sentences, which we will not discuss here). Let us see an example of it and the way *DPL* treats it.

(3) A man walks in the park. He meets a woman.

Standard first-order logic (henceforth, *PL*) translates this discourse as

$$(3a) \exists x (\text{man}(x) \ \& \ \text{walk\_in\_the\_park}(x) \\ \ \& \ \exists y (\text{woman}(y) \ \& \ \text{meet}(x, y)))^2$$

This translation reflects the anaphoric link between "he" in the second sentence and "a man" in the first by placing all the occurrences of  $x$  within the scope of  $\exists x$ . The price is that the translation is not compositional; no syntactic component of the formula translates the first sentence of (3). On the other hand,

$$(3b) \exists x (\text{man}(x) \ \& \ \text{walk\_in\_the\_park}(x)) \\ \ \& \ \exists y (\text{woman}(y) \ \& \ \text{meet}(x, y))$$

is not a correct translation, since the last occurrence of  $x$  is not bound by the quantifier  $\exists x$ . In *PL*, (3a) is not equivalent with (3b).

*DPL* offers a framework in which the syntactic scope of a quantifier occurrence (that is, the formula to which the quantifier is prefixed in the syntactic construction) separates from its semantic scope (that is, the part of the discourse in which it is able to bind variable occurrences). In *DPL*'s semantics, due to the dynamic rules governing existential quantification and conjunction, (3a) is equivalent with (3b).

Moreover, in *DPL* anaphoric linking is possible even between the conclusion of a consequence and one of its premises. Consider the following example and its attempted translations:

<sup>2</sup>As a tribute to professor Ruzsa, we are going to use his versions of the logical constants, wherever it is possible, including  $\&$  as conjunction,  $\sim$  as negation,  $\supset$  as implication,  $\equiv$  as the biconditional and  $I$  as the descriptor. See, for example, (Ruzsa 1981, 16ff.)

(4) *Witness*: A man entered the house and switched the light on. He had a knife. *Inspector*: So, he switched the light on.

(4a)  $\{\exists x (\text{man}(x) \ \& \ \text{enter\_the\_house}(x) \ \& \ \text{switch\_the\_light\_on}(x), \ \text{have\_a\_knife}(x))\} \models \text{switch\_the\_light\_on}(x)$

(4b)  $\forall x (\text{man}(x) \ \& \ \text{enter\_the\_house}(x) \ \& \ \text{switch\_the\_light\_on}(x) \ \& \ \text{have\_a\_knife}(x)) \supset \text{switch\_the\_light\_on}(x)$

This is an example that cannot be properly translated into *PI*, even if we give up compositionality. In standard semantics, (4a) is not a valid consequence, and (4b) is not even an consequence. On the other hand, *DPI* has a dynamic consequence relation that makes “cross-inferential” anaphora possible, and makes (4b) valid.

In *DPLD* we want to deal with similar examples involving atomic definite descriptions. The first one and its compositional translation are as follows:

(5) A man walks in the park. He meets a woman. The man hugs her. A man watches from a distance. He walks a dog. The dog is bored. The man is jealous.

(5a)  $\exists x \text{man}(x) \ \& \ \text{walk\_in\_the\_park}(x) \ \& \ \exists y \text{woman}(y) \ \& \ \text{meet}(x, y) \ \& \ \text{hug}(\text{Iw man}(w), y) \ \& \ \exists z \text{man}(z) \ \& \ \text{watch\_from\_a\_distance}(z) \ \& \ \exists v \text{dog}(v) \ \& \ \text{walk}(z, w) \ \& \ \text{bored}(\text{Iw dog}(w)) \ \& \ \text{jealous}(\text{Iw man}(w))$

In this discourse, all occurrences of descriptions refer back anaphorically to indefinite ones. One particularly interesting point is that the two occurrences of “the man” have two different antecedents, both of which are occurrences of the same expression “a man”. Let us call  $d_1$  and  $d_2$  the discourse referents introduced by the first and the second occurrences of  $\exists x$ , respectively; and let us call  $s_1$  the salience information that the predicate *man* is associated with  $d_1$ , and  $s_2$  the salience information that the predicate *man* is associated with  $d_2$ . Now, *DPLD* has to interpret this formula in such a way that  $s_1$  is passed to the first occurrence of *Iw man*( $w$ ), but it is replaced with  $s_2$  in the course of the updating process before the second occurrence of *Iw man*( $w$ ).

*DPLD* also has to treat descriptions that lack possible antecedents, like Lewis’s

(6) The cat is in the carton. She is asleep.

(6a)  $\text{in\_the\_carton}(\text{I}x \text{cat}(x)) \ \& \ \text{asleep}(x)$

If we meet an atomic description at the beginning of a discourse, there are two possible scenarios for evaluation. One is that—like Lewis suggests—we have salience information from some external sources. This means that the evaluation of a discourse may not start with an empty discourse information state. The other possibility is that, for some reason or another, we lack the salience information needed; for example, this is the opening sentence of a novel. In this case, the description is most likely to introduce a new discourse referent, behaving the same way as existential quantification. In either of the scenarios, the occurrence of  $x$  in the second formula is bound by the description operator in the first.

We have to deal with “cross-inferential” anaphora, too:

(7) *Witness*: A woman entered the house. She switched the light on. A man waited outside.

*Inspector*: So, the one that switched the light on was a woman.

(7a)  $(\exists x \text{ woman}(x) \ \& \ \text{enter\_the\_house}(x), \text{switch\_the\_light\_on}(x),$   
 $\exists y \text{ man}(y) \ \& \ \text{wait\_outside}(x)) \models$   
 $\text{woman}(Iz \ \text{switch\_the\_light\_on}(z))$

As the example suggests, *DPLD*'s consequence relation has to make it possible for a description in its conclusion to be bound by an existential quantifier in one of its premises. The example has another peculiarity that deserves attention. The description  $Iz \ \text{switch\_the\_light\_on}(z)$  uses a predicate that was associated with the discourse referent in one sentence later than it was introduced. This is another phenomenon in the dynamics of salience ranking to be taken account of.

Finally, let us consider the behavior of a description with respect to negated sentences.

(8) John doesn't own either a car or a motorcycle. ??The car is too expensive for him.

(8a)  $\sim ((\exists x (\text{car} \ \& \ \text{own}(\text{John}, x)) \vee \exists y (\text{motorcycle} \ \& \ \text{own}(\text{John}, x)))$   
 $\ \& \ \text{too\_expensive}(Ix \ \text{car}(x), \text{John}))$

Clearly, the description “the car” in the second sentence cannot refer back to the indefinite expression “a car” in the first. Instead, if the definite description is acceptable here at all, it introduces a new discourse referent. On the other hand, an atomic description in a negated sentence seems to be able to bind later variable occurrences:

(9) It is not the case that the cat is in the carton. She is in the garden.

(9a)  $\sim \text{in\_the\_carton}(Ix \ \text{cat}(x)) \ \& \ \text{in\_the\_garden}(x)$

Even if the referent of “the cat” is determined by external conditions, the pronoun “she” in the second sentence refers back to the description. Thus, negation seems to be externally dynamic with respect to atomic descriptions. This is essentially different from the analogous case with existential quantification:

(9') It is not the case that there is a cat in the carton. ??She is in the garden.

(9a')  $\sim (\exists x \ \text{cat}(x) \ \& \ \text{in\_the\_carton}(x)) \ \& \ \text{in\_the\_garden}(x)$

Here, the pronoun “she” appears to behave deictically, and thus the corresponding occurrence of  $x$  in the second formula is best interpreted as free.

We choose not to take this last phenomenon into account in *DPLD*. The reason is simple; no proper dynamic version of negation is known in a first-order framework. This is a fundamental problem of dynamic semantics, already observed in (Groenendijk & *al.* 1991, 99ff.), and to all appearances, it has not been solved yet.

## 3 INTERLUDE: THE DYNAMICS OF CLASSICAL DESCRIPTIONS

After making clear what we are interested in, a quick word on what we are not interested in. There is another way of using dynamic semantics in the analysis of definite descriptions; it concerns classical descriptions instead of Lewis's anaphoric ones. It is well-known that standard first-order logic cannot deal with Russell's analysis of descriptions in a compositional way.<sup>3</sup> It is less well-known that dynamic predicate logic solves this problem. This is an obvious consequence of its refined manner of handling anaphoric relations.

(10) The present king of France is bald. He shaves himself.

(10a)  $\exists x (\forall y (\text{present\_king\_of\_France}(y) \equiv y = x)$   
 $\& \text{bald}(y) \& \text{shave}(x, x))$

(10b)  $\exists x \forall y (\text{present\_king\_of\_France}(y) \equiv y = x)$   
 $\& \text{bald}(y) \& \text{shave}(x, x)$

While "the present king of France" is a syntactic component in (10), its first-order translation,  $\exists x \forall y (\text{present\_king\_of\_France}(y) \equiv y = x)$  is not a component in (10a). Nevertheless, it is present in the formula (10b), and in *DPL*. (10a) is equivalent with (10b). Similarly, the translation of the first sentence is a component of (10b), but not of (10a). However, all occurrences of  $x$  in  $\text{bald}(y)$  and  $\text{shave}(x, x)$  are anaphorically linked to  $\exists x$ . Thus, unlike *PL*, *DPL* offers a proper treatment of Russellian definite descriptions. Compositionality also makes the introduction of a classical description operator straightforward, either with a Russellian or with a Strawsonian semantics, and thus makes the semantic modelling of the dynamics of existence and uniqueness conditions straightforward. This idea is developed in (Eijck 1993), but it has little in common with our project.

4 THE SYNTAX AND SEMANTICS OF *DPL* AND *DPLR*

We are going to present dynamic predicate logic in two versions, which have the same syntax, but differ in their semantics. One of them is the well-known *DPL* presented in (Groenendijk & al. 1991). The second version, which we refer to as *DPLR*, differs from the original one at one important point: the use of referent systems. Referent systems were introduced in the semantics in (Groenendijk & al. 1996), along with the concept of a *peg*. (In fact, *DPLR* is the extensional part of the system presented in (Groenendijk & al. 1996).) Peps are the technical equivalents of discourse referents; that is, abstract entities which are introduced by indefinite descriptions like "a man", and to which discourse information is attributed instead of being attributed to the variables themselves,

<sup>3</sup>Cf. e.g. (Gamut 1991, 164.)



or the elements of the domain of discourse. We are going to refer to them as discourse referents instead of pegs.<sup>4</sup>

### Syntax

The syntax of *DPL* and *DPLR* is identical to that of standard predicate logic with identity. For simplicity, the only function symbols allowed are individual constants. Individual terms (henceforth, terms) are either variables or constants. Besides these, we have  $n$ -ary predicates in the non-logical vocabulary. Logical primitives are  $\sim$ ,  $\&$ ,  $\exists$ ,  $=$ ,  $(, )$ ; the rest of the logical symbols are defined in terms of these.

- i  $\varphi \vee \psi \iff_d \sim(\sim\varphi \& \sim\psi)$
- ii  $\varphi \supset \psi \iff_d \sim(\varphi \& \sim\psi)$
- iii  $\varphi \equiv \psi \iff_d (\varphi \supset \psi) \& (\psi \supset \varphi)$
- iv  $\forall x \varphi(x) \iff_d \sim \exists x \sim \varphi$

Now, we define formulas as usual:

- (1) If  $P$  is an  $n$ -place predicate and  $t_1, \dots, t_n$  are terms, then  $P(t_1, \dots, t_n)$  is an atomic formula.
- (2) If  $t_1$  and  $t_2$  are terms, then  $t_1 = t_2$  is an atomic formula.
- (3) Nothing else is an atomic formula. Atomic formulas are formulas.
- (4) If  $\varphi$  and  $\psi$  are formulas, then  $(\varphi \& \psi)$  is a formula.
- (5) If  $\varphi$  is a formula, then  $\sim \varphi$  is a formula.
- (6) If  $\varphi$  is a formula and  $x$  is a variable, then  $\exists x \varphi$  is a formula.
- (7) Nothing else is a formula.

### The semantics of *DPL*.

Although the original version of *DPL* is well-known, we reintroduce it in a nutshell. Dynamic semantics is semantics of information state updating. Discourse information is based on an ordinary first-order structure  $\mathcal{M} = \langle U, \varrho \rangle$ ,  $U$  being the universe of discourse and  $\varrho$  being the interpretation function. If  $a$  is an individual constant and  $P$  is an  $n$ -place predicate, then  $\varrho(a) \in U$  and  $\varrho(P) \subseteq U^n$ . This structure is not updated in the process of evaluating a discourse; it is not part of the discourse information.

We will use cylindrification as a basic operation on assignment sets, defined as

$$V[x] =_d \{v : \text{for some } v', v' \in V \text{ and } v[x]v'\},$$

<sup>4</sup>The term *discourse referent* was coined in (Karttunen 1975), along with the idea that discourse referents should be identified with natural numbers. We find this expression more expressive and less idiosyncratic than *peg*.

where

$$v[x]v' \iff_d \text{ for all variables } y \text{ different from } x, v'(y) = v(y).$$

The value  $|t|_{\mathcal{M},V}^{DPL}$  of a term  $t$  is  $\varrho(t)$  if it is a constant, and  $v(t)$  if it is a variable. Thus, the evaluation function  $\llbracket \varphi \rrbracket_{\mathcal{M}}^{DPL}$  of formulas updates information states as follows (we omit the upper index wherever it is not confusing):<sup>5</sup>

- (1)  $\llbracket P(t_1, \dots, t_n) \rrbracket_{\mathcal{M}}(V) =_d \{v \in V : \langle |t_1|_{\mathcal{M},V}, \dots, |t_n|_{\mathcal{M},V} \rangle \in \varrho(P)\}$ ;
- (2)  $\llbracket t_1 = t_2 \rrbracket_{\mathcal{M}}(V) =_d \{v \in V : \langle |t_1|_{\mathcal{M},V}, |t_2|_{\mathcal{M},V} \rangle \in \varrho(=)\}$ ;
- (3)  $\llbracket (\varphi \ \& \ \psi) \rrbracket_{\mathcal{M}}(V) =_d \llbracket (\psi) \rrbracket_{\mathcal{M}} \circ \llbracket (\varphi) \rrbracket_{\mathcal{M}}(V)$ ;
- (4)  $\llbracket \sim \varphi \rrbracket_{\mathcal{M}}(V) =_d \{v \in V : \llbracket \varphi \rrbracket_{\mathcal{M}}(\{v\}) = \emptyset\}$ ;
- (5)  $\llbracket \exists x \varphi \rrbracket_{\mathcal{M}}(V) =_d \llbracket \varphi \rrbracket_{\mathcal{M}}(V[x])$ .

Finally, a dynamic consequence relation  $\langle \varphi_1, \dots, \varphi_n \rangle \models_{DPL} \psi$  is defined as

$$\langle \varphi_1, \dots, \varphi_n \rangle \models_{DPL} \psi \iff_d \text{ for all } \mathcal{M} \text{ and } V,$$

$$\text{if } \llbracket (\varphi_n) \rrbracket_{\mathcal{M}} \circ \dots \circ \llbracket (\varphi_1) \rrbracket_{\mathcal{M}}(V) \neq \emptyset,$$

$$\text{then } \llbracket (\varphi_n) \rrbracket_{\mathcal{M}} \circ \llbracket (\varphi_n) \rrbracket_{\mathcal{M}} \circ \dots \circ \llbracket (\varphi_1) \rrbracket_{\mathcal{M}}(V) \neq \emptyset.$$

That is,  $\langle \varphi_1, \dots, \varphi_n \rangle \models_{DPL} \psi$  is valid iff whenever the evaluation process of the discourse  $\langle \varphi_1, \dots, \varphi_n \rangle$  does not result in an empty set of assignments, neither does the evaluation of  $\langle \varphi_1, \dots, \varphi_n, \psi \rangle$ . Note that, unlike in  $PL$ , the premises of a consequence relation form an ordered sequence.

### *The semantics of DPLR*

*DPLR* is a variant of *DPL* that applies a more detailed model of discourse information. In *DPL*, an information state is a set of assignments, and assignments are total functions from the set of variables to the domain of discourse. In this semantic framework discourse referents do not play an explicit role. *DPLR* differs from *DPL* in making the use of discourse referents transparent, by means of referent systems.

Discourse information is represented in *DPLR* as a triple  $\mathcal{I} = \langle d, r, V \rangle$ .  $d$  is a natural number, and serves as the set of the discourse referents that are in use at a certain point of a discourse. As it is standard in set theory, numbers are identified with the sets of their predecessors, that is,  $0 = \emptyset$  and  $n = \{0, \dots, n-1\}$ . The elements of  $d$  serve as discourse referents in a given information state.  $r$  is a partial function from the set of variables to  $n$ . The pair  $\langle d, r \rangle$  is called a referent

<sup>5</sup>Instead of updating a set of assignments as a whole, (Groenendijk & al. 1991) updates each assignment separately. Our version is standard in dynamic systems that give an explicit definition of information state. We first saw *DPL* presented this way in (Kálmán & al. 2001, 62ff.)

system.  $V$  is a set of evaluation functions from  $d$  to  $U$ . Thus, if a variable  $x$  has a value, then its value is given by  $v(r(x))$  for each  $v \in V$ . But it is not the case that every variable has a value; in fact, the evaluation of any discourse only necessitates that a finite number of them has. A variable that does not occur in a discourse does not have a value. Discourse information is treated in a dynamic way; in the the course of evaluating a discourse, each occurrence of a term updates the actual referent system, the same way as each occurrence of a formula does.

Some technical concepts will be useful in the definition of semantics. The first is that of updating of a referent system with a new variable. This will be used in the evaluation of quantified formulas.

$$\langle d, r \rangle [x] =_d \langle d + 1, (r \setminus \{ \langle x, i \rangle : i < d \}) \cup \{ \langle x, d \rangle \} \rangle$$

That is, a new discourse referent is introduced into the referent system as the value of the variable  $x$ . Meanwhile, if  $x$  already had a value, the old value is deleted. A set of assignments  $V$  is updated with a new discourse referent in a similar way:

$$V[d] =_d \{ v : (\exists v' \in V) (\exists u \in U) v = v' \cup \{ \langle d, u \rangle \} \}$$

(Note that unlike variables, discourse referents are not reused in the evaluation process; and although we don't rule out the possibility that different variables are associated with the same discourse referent, nothing in our semantics enforces such a situation.)

To make the following definitions more transparent, we will follow a simple notational rule for information states and first-order structures:  $I = \langle d, r, V \rangle$ ,  $I' = \langle d', r', V' \rangle$ ,  $I'' = \langle d'', r'', V'' \rangle$  etc; and  $\mathcal{M} = \langle U, \varrho \rangle$ ,  $\mathcal{M}' = \langle U', \varrho' \rangle$  etc. This way it will be easy to identify the semantic components.

A discourse information state  $\mathcal{I}$  is updated by updating both  $\langle d, r \rangle$  and  $V$ :

$$\mathcal{I}[x] =_d \langle \langle d, r \rangle [x], V[d] \rangle^6$$

We call a discourse information state *empty* iff it has an empty set of assignments. An empty information state may or may not have an empty referent system.

The value  $|t|_{\mathcal{M}, \mathcal{I}}^{DPIR}$  of a term  $t$  is  $\varrho(t)$  if it is a constant, and  $v(r(t))$  if it is a variable with a previous occurrence. Variables with no previous occurrences have no value. Thus, the evaluation function  $\llbracket \varphi \rrbracket_{\mathcal{M}}^{DPIR}$  of formulas is essentially partial. (We omit the upper indices wherever it is not confusing.) Formulas with semantically free variables have no semantic value. The semantic value of a formula, if it exists, is a function from discourse information states to discourse information states, defined in the following clauses:

<sup>6</sup>As is customary, for convenience, we identify  $\langle \langle a, b \rangle, c \rangle$  with  $\langle a, \langle b, c \rangle \rangle$ . Also, we identify  $\langle a \rangle$  with  $a$ .

- (1)  $\llbracket P(t_1, \dots, t_n) \rrbracket_{\mathcal{M}}(\mathcal{I}) =_d \langle d, r, \{v \in V : \langle |t_1|_{\mathcal{M}, \mathcal{I}}, \dots, |t_n|_{\mathcal{M}, \mathcal{I}} \rangle \in \varrho(P) \} \rangle$ ;
- (2)  $\llbracket t_1 = t_2 \rrbracket_{\mathcal{M}}(\mathcal{I}) =_d \langle d, r, \{v \in V : \langle |t_1|_{\mathcal{M}, \mathcal{I}} = |t_2|_{\mathcal{M}, \mathcal{I}} \rangle \} \rangle$ ;
- (3)  $\llbracket (\varphi \ \& \ \psi) \rrbracket_{\mathcal{M}}(\mathcal{I}) =_d \llbracket (\psi) \rrbracket_{\mathcal{M}} \circ \llbracket (\varphi) \rrbracket_{\mathcal{M}}(\mathcal{I})$ ;
- (4)  $\llbracket \sim \varphi \rrbracket_{\mathcal{M}}(\mathcal{I}) =_d \langle d, r, \{v \in V : \llbracket \varphi \rrbracket_{\mathcal{M}}(\langle d, r, \{v\} \rangle) \text{ is empty} \} \rangle$ ;
- (5)  $\llbracket \exists x \varphi \rrbracket_{\mathcal{M}}(\mathcal{I}) =_d \llbracket \varphi \rrbracket_{\mathcal{M}}(\mathcal{I}[x])$ .

Finally, the definition of dynamic consequence is as follows:

$\langle \varphi_1, \dots, \varphi_n \rangle \models \psi$  iff for all  $\mathcal{M}$  and  $\mathcal{I}$ , if

$$\llbracket \psi \rrbracket_{\mathcal{M}} \circ \llbracket \varphi_n \rrbracket_{\mathcal{M}} \circ \dots \circ \llbracket \varphi_1 \rrbracket_{\mathcal{M}}(\mathcal{I})$$

exists and

$$\llbracket \varphi_n \rrbracket_{\mathcal{M}} \circ \dots \circ \llbracket \varphi_1 \rrbracket_{\mathcal{M}}(\mathcal{I})$$

is nonempty, then

$$\llbracket \psi \rrbracket_{\mathcal{M}} \circ \llbracket \varphi_n \rrbracket_{\mathcal{M}} \circ \dots \circ \llbracket \varphi_1 \rrbracket_{\mathcal{M}}(\mathcal{I})$$

is nonempty.

### *The problem of partiality*

We have seen how *DPIR* makes the semantic role of a discourse referent explicit. The price of this is partiality. In *DPIR*, a semantically free variable occurrence—that is, an occurrence of a variable that has not been associated with a discourse referent—results in a semantic value gap. This is not necessarily a problem. As it turns out from the above definitions, the evaluation process of a discourse  $\langle \varphi_1, \dots, \varphi_n \rangle$  does not necessarily start with an empty referent system. If a variable  $x$  occurs in a discourse without an antecedent  $\exists x$ , it may still have a value, provided by the initial information state. This way *DPIR* gives an account of the deictic use of pronouns of natural language.

However, a natural language discourse does not always come to a halt when a new pronoun is introduced without salient discourse referent. It is more likely that a new discourse referent is tacitly introduced. This resembles the case of using an atomic description without any salient referent, like in example (6) above. For this reason, in *DPLD* we do not follow (Groenendijk & *al.* 1996) in making semantic rules partial.

## 5 THE SYNTAX AND SEMANTICS OF *DPLD*

In this section we introduce *DPLD*, a system of dynamic predicate logic with atomic descriptions, as a modified version of *DPIR*. The main differences are the use of the descriptor in the syntax, and salience ranking as a component of discourse information in semantics.

### Syntax

In all our examples, we used one-place predicates in the descriptions. We restrict ourselves to them in the syntax, because we do not intend to model the dynamics of embedded and open descriptions, like  $\text{Ix } R(x, \text{Iy } P(y))$  (“the man’s friend”) and  $\text{Ix } R(x, y)$  (“his friend”). Adapting our semantic rules to atomic descriptions with many-place predicates is straightforward but slightly complicated.

We define terms by the following clauses:

- (1) Variables and individual constants are terms.
- (2) If  $P$  is a one-place predicate and  $x$  is a variable, then  $\text{Ix } Px$  is a term.

The rest of the syntax remains intact.

### Semantics

As in *DPLR*, the definition of semantics begins with the concept of a discourse information state. We use referent systems and assignment sets as in *DPLR* (but as we will see, they are updated differently). We enrich discourse information states with a new component, salience ranking;  $\mathcal{I} = \langle d, r, S, V \rangle$ , where  $S$  is a salience ranking.

A salience ranking is an ordered tuple of salience information bits. Salience bits, on their turn, are ordered pairs. If  $P$  is a one-place predicate and  $d$  is a discourse referent, then  $\langle P, d \rangle$  is a salience bit. (Note that salience bits have both syntactic and semantic elements.) Let us explain its use through simple examples.

Let  $P$  be a one-place predicate,  $d$  a discourse referent, and let the salience information bit  $s = \langle P, d \rangle$  be at the head of our actual salience ranking. Now, if a description  $\text{Ix } Px$  occurs in the course of evaluation, then our actual referent system  $\langle d', r \rangle$  is updated with the pair  $\langle x, d \rangle$ . On the other hand, if we find a number of bits of the form  $s = \langle P, d \rangle$  in our actual salience ranking, then the referent system is updated with the discourse referent in the leftmost bit. And finally, if there is no bit of the form  $\langle P, d \rangle$  on the list, then  $x$  is associated with a new discourse referent, just like in the case of *DPL*’s existential quantification.

The formal definition of updating a referent system is

$$\langle d, r \rangle[x/d'] =_d \langle \max(d, d' + 1), r \setminus \{ \langle x, i \rangle : i < d \} \cup \{ \langle x, d' \rangle \} \rangle$$

The difference between this update definition and the one given in the last section is that  $x$  is not necessarily associated with a new discourse referent. As a consequence, different variables can be associated with the same discourse referent. An occurrence of  $\text{Ix } P(x)$  can refer back to  $\exists y P(y)$  in this way; although  $x$  is not bound by  $\exists y$ ,  $x$  and  $y$  are associated with the same discourse referent, and hence they have the same value.

Let, again,  $P$  be a one-place predicate, and  $t$  a term associated with the discourse referent  $d$ ; that is,  $r(t) = d$ . Let the atomic formula  $P(t)$  occur at a certain point of the discourse. Then the salience ranking  $S$  is updated with a new bit  $\langle P, d \rangle$ ; that is, this pair is attached to the head of the ranking:

$$S[P(t)] =_d \langle \langle P, r(t) \rangle, S \rangle$$

Since the last information bit is always the leftmost one, the discourse referent this bit offers is more salient than its rivals that are offered further down in the ranking. In simple terms, the last thing mentioned is the most salient.

It will be useful to define a function  $S(P)$  that gives back the most salient discourse referent associated with a predicate  $P$  in a given salience ranking  $S$ . If there is no salient discourse referent, then the value of  $S(P)$  is  $P$ . We define  $S(P)$  recursively with respect to the length of  $S$ .

- (1) If  $S = \emptyset$ , then  $S(P) = P$ ;
- (2) if  $S = \langle \langle P, d \rangle, S' \rangle$  for some  $d$  and  $S'$ , then  $S(P) = d$ ;
- (3) if  $S = \langle \langle Q, d \rangle, S' \rangle$  for some  $Q$  that is different from  $P$ ,  $d$  and  $S'$ , then  $S(P) = S'(P)$ .

The last component of a discourse information state is a set  $V$  of assignments. It is updated with a discourse referent the same way as in *DPL*. There are two ways of updating the discourse information state with a new variable  $x$ . In the first case,  $x$  is associated with an existing discourse referent:

$$\mathcal{I}[x/d'] =_d \langle \langle d, r \rangle [x/d'], S, V \rangle$$

In the second case, a new discourse referent is introduced along with a new variable:

$$\mathcal{I}[x] =_d \langle \langle d, r \rangle [x/d], S, V[d] \rangle$$

We have seen that terms are capable of updating discourse information. Now we define this update function  $\llbracket t \rrbracket_{\mathcal{M}}^{DPL,D}$ .

- (1)  $\llbracket x \rrbracket_{\mathcal{M}}(\mathcal{I}) =_d \begin{cases} \mathcal{I}[x] & \text{if } x \notin \text{dom}(r); \\ \mathcal{I} & \text{if } x \in \text{dom}(r); \end{cases}$
- (2)  $\llbracket a \rrbracket_{\mathcal{M}}(\mathcal{I}) =_d \mathcal{I}$ ;
- (3)  $\llbracket \text{I}x P(x) \rrbracket_{\mathcal{M}}(\mathcal{I}) =_d \begin{cases} \mathcal{I}[x] & \text{if } S(P) = P; \\ \mathcal{I}[x/d'] & \text{if } S(P) = d'. \end{cases}$

That is, an occurrence of a variable  $x$  that has not yet been associated with a discourse referent behaves like an existential quantifier in the sense that it introduces a new discourse referent. Occurrences of  $x$  that are already associated with a discourse referent leave discourse information unchanged. Individual constants do not change discourse information either. An occurrence of an atomic description  $\text{I}x P(x)$  for which there is no salient referent introduces

a new one. Otherwise the  $x$  of  $Ix P(x)$  is associated with the most salient discourse referent.

The value  $|t|_{\mathcal{M}, \mathcal{I}}^{DPLD}$  of a term  $t$  is defined as follows:

- (1)  $|a|_{\mathcal{M}, \mathcal{I}} =_d \varrho(t)$ ;
- (2)  $|x|_{\mathcal{M}, \mathcal{I}} =_d v(r(x))$ ;
- (3)  $|ixP(x)|_{\mathcal{M}, \mathcal{I}} =_d v(r(x))$ .

Thus every occurrence of any term has a value in *DPLD*. Unlike *DPLR*, the semantics of *DPLD* is not a partial one.

Now we are ready to evaluate formulas. The information update function  $[\![\varphi]\!]_{\mathcal{M}}^{DPLD}$  is defined in the following clauses:

- (1)  $[\![P(t)]\!]_{\mathcal{M}}(\mathcal{I}) =_d \langle d', r', S'[P(t)], \{v \in V' : |t|_{\mathcal{M}, \mathcal{I}} \in \varrho(P)\} \rangle$ ,  
where  $\langle d', r', S', V' \rangle = [\![t]\!]_{\mathcal{M}}(\mathcal{I})$ ;
- (2)  $[\![P(t_1, \dots, t_n)]\!]_{\mathcal{M}}(\mathcal{I}) =_d \langle d', r', S', \{v \in V' : \langle |t_1|_{\mathcal{M}, \mathcal{I}}, \dots, |t_n|_{\mathcal{M}, \mathcal{I}} \rangle \in \varrho(P) \rangle \rangle$ ,  
where  $n > 1$  and  $\langle d', r', S', V' \rangle = [\![t_n]\!]_{\mathcal{M}} \circ \dots \circ [\![t_1]\!]_{\mathcal{M}}(\mathcal{I})$ ;
- (3)  $[\![t_1 = t_2]\!]_{\mathcal{M}}(\mathcal{I}) =_d \langle d', r', S', \{v \in V' : \langle |t_1|_{\mathcal{M}, \mathcal{I}} = |t_2|_{\mathcal{M}, \mathcal{I}} \rangle \rangle \rangle$ ,  
where  $\langle d', r', S', V' \rangle = [\![t_2]\!]_{\mathcal{M}} \circ [\![t_1]\!]_{\mathcal{M}}(\mathcal{I})$ ;
- (4)  $[\![\varphi \& \psi]\!]_{\mathcal{M}}(\mathcal{I}) =_d [\![\psi]\!]_{\mathcal{M}} \circ [\![\varphi]\!]_{\mathcal{M}}(\mathcal{I})$ ;
- (5)  $[\![\sim \varphi]\!]_{\mathcal{M}}(\mathcal{I}) =_d \langle d, r, S, \{v \in V : [\![\varphi]\!]_{\mathcal{M}}(\langle d, r, S, \{v \} \rangle) \text{ is empty} \} \rangle$ ;
- (6)  $[\![\exists x \varphi]\!]_{\mathcal{M}}(\mathcal{I}) =_d [\![\varphi]\!]_{\mathcal{M}}(\mathcal{I}[x])$ .

Thus, the definition deviates from the one of *DPL* only in the evaluation of atomic formulas. Dynamic inference is defined the same way as in *DPL*.

To see how discourse information updating works, it will be instructive to see the evaluation of a particular formula. Let us consider the first three sentences of example (5). To make the formula shorter, we abbreviate the predicates *man*, *walk\_in\_the\_park*, *woman*, *meet* and *hug* as  $P$ ,  $Q$ ,  $T$ ,  $R$  and  $S$ , respectively.

(5') A man walks in the park. He meets a woman. The man hugs her.

(5a')  $\exists x P(x) \& Q(x) \& \exists y T(y) \& R(x, y) \& S(Iw P(w), y)$

We present the evaluation process of the formula in an informal fashion.

- (1) We start with an empty referent system, an empty salience ranking and an empty assignment set. That is, we do not make use of an external context.
- (2)  $\exists x P(x)$  introduces a new discourse referent, and associates  $x$  with it. Salience information associates the new discourse referent with the predicate  $P$ .  $d_1 = 1$ ;  $r_1 = \{\langle x, 0 \rangle\}$ ;  $S_1 = \langle P, 0 \rangle$ ;  $V_1 = \{\{\langle 0, u \rangle\} : u \in \varrho(P)\}$ .
- (3)  $Q(x)$  updates salience information.  $d_1 = d_1$ ;  $r_2 = r_1$ ;  $S_2 = \langle \langle Q, 0 \rangle, \langle P, 0 \rangle \rangle$ ;  $V_1 = \{\{\langle 0, u \rangle\} : u \in \varrho(P), u \in \varrho(Q)\}$ .
- (4)  $\exists y T(y)$  introduces a new discourse referent, associates  $y$  with it, and updates salience information.  $d_3 = 2$ ;  $r_3 = \{\langle x, 0 \rangle, \langle y, 1 \rangle\}$ ;

- $S_3 = \langle \langle T, 1 \rangle, \langle Q, 0 \rangle, \langle P, 0 \rangle \rangle$ ;  $V_3 = \{ \{ \langle 0, u \rangle, \langle 1, u' \rangle \} : u \in \varrho(P), u \in \varrho(Q), u' \in \varrho(T) \}$ .
- (5)  $R(x, y)$  does not change either the referent system or the salience ranking.  $d_4 = d_3$ ;  $r_4 = r_3$ ;  $S_4 = S_3$ ;  $V_4 = \{ \{ \langle 0, u \rangle, \langle 1, u' \rangle \} : u \in \varrho(P), u \in \varrho(Q), u' \in \varrho(T), \langle u, u' \rangle \in \varrho(R) \}$ .
- (6)  $Iw P(w)$  introduces a new variable in the referent system. Since  $P$  occurs only once in  $S_4$ , this occurrence determines the discourse referent associated with  $w$ :  $r_5(w) = S_4(P) = 0$ . Thus,  $d_5 = d_4$ ;  $r_5 = \{ \langle x, 0 \rangle, \langle y, 1 \rangle, \langle w, 0 \rangle \}$ ;  $S_5 = S_4$ ;  $V_5 = V_4$ .
- (7) Finally,  $S(Iw P(w), y)$  does not change either the referent system or the salience ranking.  $d_6 = d_5$ ;  $r_6 = r_5$ ;  $S_6 = S_5$ ;  $V_6 = \{ \{ \langle 0, u \rangle, \langle 1, u' \rangle \} : u \in \varrho(P), u \in \varrho(Q), u' \in \varrho(T), \langle u, u' \rangle \in \varrho(R), \langle u, u' \rangle \in \varrho(S) \}$ .

## REFERENCES

- Eijck, J. van, 1993, The dynamics of description. *Journal of Semantics* 10, 223–267.
- Eijck, J. van and Kamp, H., 1997, Representing discourse in context. In J. van Benthem and A. ter Meulen (eds.), *Handbook of Logic and Language*. Elsevier, Amsterdam & MIT Press, Cambridge (MA), 179–237.
- Gamut, L. T. F., 1991, *Logic, Language, and Meaning*. University of Chicago Press, Chicago. (L. T. F. Gamut is a collective pseudonym for J. van Benthem, J. Groenendijk, D. de Jongh, M. Stokhof and H. Verkuyl.)
- Groenendijk, J. and Stokhof, M., 1991, Dynamic predicate logic. *Linguistics and Philosophy* 14, 39–100.
- Groenendijk, J., Stokhof, M. and Veltman, F., 1996, Coreference and modality. In S. Lappin (ed.), *The Handbook of Contemporary Semantic Theory*, Blackwell, 179–214.
- Kálmán, L. and Rádai, G., 2001, *Dinamikus szemantika (Dynamic Semantics)*; in Hungarian). Budapest, Osiris.
- Karttunen, L., 1975, Discourse referents. In J. McCawley (ed.), *Syntax and Semantics 7: Notes from the Linguistic Underground*. Academic Press, New York, 363–385.
- Lewis, D., 1979, Scorekeeping in a language game. *Journal of Philosophical Logic* 8, 339–359.
- Ruzsa, I., 1981, *Modal Logic with Descriptions*. Martinus Nijhoff, De Hague.



# PHILOSOPHY OF MATHEMATICS

ZOLTÁN GENDLER SZABÓ

## Tasks and Ultra-tasks

**Abstract.** Can we count the primes? There is a near unanimous consensus that in principle we can. I believe the near-consensus rests on a mistake: we tend to confuse counting the primes with counting each prime. To count the primes, I suggest, is to come up with an answer to the question “How many primes are there?” because of counting each prime. This, in turn requires some sort of dependence of outcome on process. Building on some ideas from Max Black, I argue that—barring very odd laws of nature—such dependence cannot obtain.

### 1 COUNTING THE PRIMES

There are countably many primes but this does not settle the question whether the primes can be counted. “Countably infinite” is a technical term which applies to multitudes equinumerous with the natural numbers. The question I want to pursue here is whether any such multitude can be counted. I will talk about counting the primes but that is just an example—I could talk about counting any countably infinite set.

The obvious difficulty with counting the primes is that even if you could go on forever, it seems there would always remain more to count. But what if you could count faster and faster as you proceed? Then, we are taught, there would be no problem. You could call the first prime within the first 1 minute, the second within the next 1/2 minute, the third within the next 1/4 minute, ... the  $n$ th within the next  $1/2^{n-1}$  minute, ... and so on. After 2 minutes you would have counted *each prime*. But would you thereby have counted *the primes*?

There is a difference between counting each prime and counting the primes. If you count the natural numbers you have counted each prime but you have not counted the primes. In general, if all  $F$ s are  $G$ s, then in counting the  $G$ s you count each  $F$  but not necessarily the  $F$ s. Should we perhaps say that counting the  $F$ s is counting each  $F$  without counting *non- $F$ s*? This won't do, as the following

example illustrates. Suppose there is a birthday party at your house for your six year old and you ask him to count his guests. He counts them in the living room, in the dining room, in the kitchen, and so on for all the rooms of the house. He is lucky—the children don't move around and he doesn't miss any. He is also smart—knowing that he is not too good with large numbers, whenever he is done with a room he writes down the result, moves to the next room, and starts the count from 1 again. Now suppose you asked him after all the numbers are written down but before they are added up “Did you count the guests?” The natural response would be “Not yet.” Each and only the guests have been counted but the guests have not been counted.

Why not? Because the child doesn't know how many guests there are, one might think. But this is not exactly right: coming to know the number of guests is neither necessary nor sufficient for counting them. If the child makes a mistake in adding up the numbers on the paper he will have counted the guests without coming to know their number; if he is told how many guests there are before he adds up the numbers on the paper he will know the number of guests without having counted them. To count the *F*s, I suggest, is to come up with some sort of (perhaps incorrect, perhaps poorly justified) answer to the question “How many *F*s are there?” *as a result of* counting each *F*.<sup>1</sup> Counting each guest in the usual way *leads to* an answer to the question “How many guests are there?”; counting each guest the way the child did it does not.

I believe the near-consensus that the primes could in principle be counted rests on a mistake: we tend to confuse counting the primes with counting each prime. It may be obvious that we can count each prime in the exponentially accelerating way described above but it is not obvious that this process leads to an answer to the question “How many primes are there?”. When we count the primes below 1000 in the usual way, at each step we come to have an answer to the question “How many primes have you already counted?” When we reach what we take to be the largest prime below 1000 we *ipso facto* arrive at an answer to the question “How many primes are there below 1000?” But when we count each prime there is no last step in the count. Accordingly, it is anything but obvious that we can come to have an answer to the question “How many primes are there?” simply as a result of counting each prime one after the other.

The expressions “as a result of” and “leads to” are not among the clearest in our philosophical vocabulary. If I had an analysis of the concepts they express, I would gladly dispense with them. Unfortunately, I have no analysis and I am not optimistic about finding one anytime soon. (In section 3, I will suggest a necessary condition for a process to lead to a state.) One thing is clear: these

<sup>1</sup> This is compatible with the possibility of counting the *F*s when one already has an answer to the question “How many *F*s are there?” We can come up with an answer to a question as a result of a procedure even if we already have an answer to that question.

locutions express some sort of dependence, presumably a causal one, of outcome on process. I will argue that the outcome of an infinite count (having an answer to a certain cardinality question) cannot depend in the right way on the process of counting. If I am right, the conventional wisdom is wrong: the primes (or, for that matter, any other countably infinite multitude) cannot be counted.

## 2 BLACK'S ARGUMENT

More than half a century ago, Max Black argued that certain infinitary tasks are impossible to perform (Black 1951). When Black's article originally appeared, it produced a flurry of reactions, including Thomson 1954, where the famous lamp is discussed. (Benacerraf 1962) was seen as a decisive refutation of Black's argument, which in turn has largely faded from philosophical discussion. I think the dismissal was too quick: there is something important Black was right about, though it is not exactly what he thought he was right about.

The target of Black's criticism is the standard "mathematical" resolution of Zeno's paradox. According to this line of thought, Achilles can catch up with the tortoise by reaching the place  $p_0$  where the tortoise had started the race, then reaching the place  $p_1$  the tortoise had reached by the time Achilles reached  $p_0$ , then reaching the place  $p_2$  the tortoise had reached by the time Achilles reached  $p_1$ , ... and so on. This amounts to performing an infinite number of tasks: moving first to  $p_0$ , then from  $p_0$  to  $p_1$ , then from  $p_1$  to  $p_2$ , ... and so on. If Achilles runs faster than the tortoise, the amount of time it takes to perform all these tasks one after the other is finite. Hence, we are told, there is no problem with catching up with the tortoise in this manner.

Black disagrees. His central complaint is that the "mathematical" resolution of the paradox "tells us, correctly, when and where Achilles and the tortoise will meet, *if* they meet; but it fails to show that Zeno was wrong in claiming they *could not* meet" (Black 1951, 93). Not that anyone should doubt that Achilles can catch up with the tortoise – Black certainly does not. His question is whether Achilles can catch up with the tortoise by performing an infinite series of tasks. Black thinks there is a serious difficulty with that idea "and it does not help to be told that the tasks become easier and easier, or need progressively less and less time in the doing" (Black 1951, 94).

What is the difficulty? Black asks us to imagine a mechanical scoop with an infinitely long narrow tray to its left and another to its right. The scoop is capable of moving marbles from one tray to the other at any finite speed. Let machine *Alpha* work as follows: Initially there are infinitely many marbles in the left tray and the right tray is empty. *Alpha* moves the first marble from left to right in 1 minute and then rests for 1 minute, it moves the second marble from left to right in 1/2 minute and then rests for 1/2 minute, it moves the third marble from left

to right in  $1/4$  minute and then rests for  $1/4$  minute, ... , it moves the  $n$ th marble from left to right in  $1/2^{n-1}$  minute and then rests for  $1/2^{n-1}$  minute, ... and so on. After 4 minutes, *Alpha* stops. Black argues that it is impossible for *Alpha* to move the marbles from left to right.

Since there is nothing special about moving marbles, if Black is right about *Alpha* then all sorts of infinitary tasks are impossible to perform. Imagine that whenever *Alpha* moves a marble from left to right, Achilles traverses one of the intervals mentioned in the “mathematical” resolution to the paradox. If *Alpha* cannot move the marbles left to right by moving them one by one, then Achilles can presumably also not catch up with the tortoise by traversing the intervals one after the other. *Alpha* is a good stand-in for counting the primes as well. Imagine that a prime is written on each of the marbles and that we call a prime when *Alpha* moves the appropriate marble from left to right. If the primes can be counted at all, then presumably they can be counted in this way—at least, those who think otherwise should explain why counting the primes by calling each prime in an exponentially accelerating way is possible, but moving the marbles from one tray to another by moving each marble in the same exponentially accelerating way is not.<sup>2</sup> I will assume that if *Alpha* cannot move the marbles from left to right, then the primes cannot be counted either.

So why does Black think that *Alpha* cannot move the marbles from left to right? The argument is presented in a somewhat oblique fashion and it can be interpreted in more than one way; I will give my best and most charitable reconstruction. The first thing to note is that in order to move the marbles from left to right, *Alpha* has to bring it about that at some time, the marbles are on the right. (Suppose it is your custom to remove no more than a couple of books at a time from your library to your study. In the evening, you always return them. Still, over the years you have carried each of the hundreds of books from your library into the study at one time or other. Then saying “I have moved the books from the library to the study” would be false, given the most natural reading of this sentence. Similarly, if the marbles are never all on the right then *Alpha* did not move them to the right.<sup>3</sup>) The second thing to note is that *Alpha* cannot bring it about that the marbles are all on the right unless it brings that about

<sup>2</sup> It is not enough to point out that counting is a mental process, and moving marbles a physical one. Some dualists believe that there are mental processes fundamentally different from all physical ones. Even if they are right, it does not follow that infinite counts are among the mental processes that lack a physical analogue. It is beyond doubt that one can model finite counts with finite marble transfers—why think that we cannot model infinite counts with infinite marble transfers?

<sup>3</sup> I don’t deny that “move the marbles from left to right” has a distributive reading that is synonymous with “moving each of the marbles from left to right,” just as “counting the guests” has a distributive reading that is synonymous with “counting each of the guests.” But the dominant readings are the collective ones, and those are the ones that I am concerned with.

when it stops. (Before *Alpha* stops there are always marbles on the left. After it stops it cannot bring anything about.) It follows that *Alpha* cannot move the marbles from left to right unless it can bring it about that they are all on the right when it stops. Black seeks to show that *Alpha* cannot bring this about.

Black goes on to consider two other machines, *Beta* and *Gamma*. They are similar to *Alpha* but they share their trays and they work in tandem: Initially there is a single marble in the left tray and the right tray is empty. *Beta* moves the marble from left to right in 1 minute while *Gamma* rests, then *Gamma* moves the marble from right to left in 1 minute while *Beta* rests, then *Beta* moves the marble from left to right in  $1/2$  minute while *Gamma* rests, then *Gamma* moves the marble from right to left in  $1/2$  minute while *Beta* rests, ... , then *Beta* moves the marble from left to right in  $1/2^n$  minute while *Gamma* rests, then *Gamma* moves the marble from right to left in  $1/2^n$  minute while *Beta* rests, ... and so on. After 4 minutes, *Beta* and *Gamma* stop.

We can think of *Alpha*, *Beta* and *Gamma* as having the same kind of task: to bring it about that all the marbles they are working with are on one side when they stop. *Alpha* accomplishes its task just in case it brings it about that the infinitely many marbles that are initially on the left are all on the right when it stops, *Beta* accomplishes its task just in case it brings it about that the single marble that is initially on the left is on the right when it stops, and *Gamma* accomplishes its task just in case it brings it about that the single marble that is initially on the left (but is subsequently moved to the right by *Beta*) is on the left when it stops. Now, we can reason as follows:

- (1) Necessarily, *Alpha* accomplishes its task if and only if *Beta* does.
- (2) Necessarily, *Beta* accomplishes its task if and only if *Gamma* does.
- (3) Necessarily, either *Beta* or *Gamma* does not accomplish its task.

Therefore,

- (4) Necessarily, *Alpha* does not accomplish its task.

*Alpha*'s task was to bring it about that the marbles are on the right when it stops. Since it cannot accomplish this task, it cannot bring it about the marbles are on the right, and hence, it cannot move the marbles from left to right. This is Black's argument, as I understand it.

### 3 IN SUPPORT OF THE ARGUMENT

The argument is clearly valid; the question is whether it is sound. Black supports (1) by emphasizing that *Alpha* and *Beta* are intrinsically the same: if you put them side by side their moves would be entirely parallel. The difference is

merely that while *Alpha* transfers an infinite number of qualitatively identical marbles, *Beta* transfers the same marble an infinite number of times. Something similar can be said in defense of (2). While there certainly is an intrinsic difference between *Beta* and *Gamma* – after a move that took  $1/2^n$  minute the former rests  $1/2^n$  minute, while the latter only  $1/2^{n+1}$  minute – this seems negligible in the light of the fact the two machines go through an identical sequence of pairs of moves and rests. The difference is in the direction of the moves and in the order within the pairs: *Beta* moves first and rests afterwards, while for *Gamma* it is the other way around. Regarding (3), Black simply points out that after 4 minutes the marble must end up somewhere, so either *Beta* or *Gamma* must fail to accomplish its task.

How strong these considerations are depends on what kind of necessity is at stake. Physical necessity won't do—it makes the premises as well as the conclusion trivial. The workings of *Alpha*, *Beta*, and *Gamma* are all physically impossible: they require motion faster than the speed of light. Black is clear that he has logical necessity in mind, but that won't do either. It is not logically necessary for the marble to end up in just one of the trays—bilocation of an object is not only possible but perhaps even actual at the micro-physical level.

So Black's argument is uninteresting when the modality is construed as physical necessity and unsound when it is construed as logical necessity. But ordinarily when we wonder whether *Alpha* can accomplish its task we have neither the physical nor the logical interpretation of the modal auxiliary in mind—we wonder whether there are more or less homely possible worlds where *Alpha* succeeds. The fact that nothing moves faster than the speed of light in the actual world and the fact that marbles are at different places at the same time in some remote possible worlds are irrelevant to the question we are after. This is not a special feature of the problem at hand: in general when we are asked whether something can be done we are supposed to ignore parochial limitations and far-fetched possibilities.<sup>4</sup> I interpret the question whether *Alpha* can move the marbles from left to right as asking whether it is possible—as we ordinarily understand what is possible in the sorts of contexts set by the description of the machine—for *Alpha* to accomplish its task.

Given the ordinary understanding of necessity (3) is in good shape: if *Beta* and *Gamma* both accomplish their tasks, then the marble is both on the left and on the right, which is certainly a far-fetched possibility. But the other two premises remain problematic even under the ordinary construal of the modality. The problem is that as long as the machines operate independently of each

<sup>4</sup> Suppose someone asks you the following: "Can you swim across this river?" If you say "No, I don't have my goggles with me" you misconstrue the question by failing to ignore parochial limitations. If you say "Yes, but I would need to learn how to swim first" you misconstrue the question by failing to ignore remote possibilities. What counts as parochial limitation or remote possibility depends on the context.

other, something can interrupt the working of one but not the other. It could happen, for example, that before the full 4 minutes elapse someone crushes *Beta* with a hammer. Then *Beta* surely does not accomplish its task but we are given no reason to think that *Alpha* fails too. Alternatively, it might be that while *Gamma* is making one of its moves and *Beta* is at rest, the latter is replaced by a duplicate machine which picks up the moving of the marble where *Beta* left off. Then *Beta* does not accomplish its task but for all we know *Gamma* might. To do justice to the intuition behind Black's argument, we must ignore not only far-fetched possible worlds but also nearby ones where something interferes with the proper functioning of the machines.

Suppose we do that; then the following case can be made for (2). If the machines work uninterrupted, the case of *Beta* and *Gamma* exhibits global symmetry. If the world is not too far-fetched, we should not expect anything to break this symmetry. While the description of the case leaves it open where the marble is located after the full 4 minutes has elapsed, reasonable guesses do not privilege any particular position: the chance that the marble should end up on the left equals the chance that it should end up on the right. Suppose then that the marble ends up on the right. Then *Gamma*, the machine whose task it was to bring it about that the marble ends up on the left, has failed. But *Beta* has not succeeded either. *Beta*'s moves did not raise the chances of the marble ending up on the right above the chances of it *not* ending up there, so the moves did not *bring it about* that the marble ended up on the right.<sup>5</sup> If we assume the marble ends up on the left (or somewhere other than the left or right trays), an analogous argument shows that neither *Beta* nor *Gamma* accomplishes its task.

I think these considerations show that, if we interpret the modality appropriately, premises (2) and (3) of Black's argument come out true: in nearby worlds where their work remains uninterrupted, neither *Beta* nor *Gamma* succeeds. However, many would insist, the case of *Alpha* is different. *Beta*'s work is constantly undone by *Gamma*, and vice versa, while nothing whatsoever interferes with the work of *Alpha*. This might explain how *Beta* and *Gamma* could fail while *Alpha* succeeds. While I don't think this is correct, I agree that (1) is in need of support.

We need a way to bolster the intuition that the makeup and task of *Alpha* and *Beta* are sufficiently similar that if they both work uninterrupted, the possibilities where one succeeds without the other are far-fetched. Here is a somewhat analogous problem. Consider two pebbles *Aleph* and *Beth*, the former being heavier than the latter. Suppose we drop both at the same time from the same height. There is a powerful intuition that *Aleph* will reach the ground

<sup>5</sup> I am assuming here only a necessary (but not sufficient) condition for a process bringing about a state: if the process *P* brings about the state *S*, then the objective chance of *S* holding after *P* is higher than the objective chance of *S* not holding after *P*.

before *Beth*. To combat this intuition, Galileo invited us to consider a case where we tie *Aleph* and *Beth* together and drop them from the same height. Call the tied up object *AlephBeth*. *AlephBeth* can be seen in two different ways—as a unit or as two separate objects. If heavier objects in general fall faster than lighter ones, *AlphaBeth* as a unit would have to fall faster than *Aleph* (since its weight exceeds the weight on *Aleph*) but *AlephBeth* as two separate objects would have to fall slower than *Aleph* (since it is held back by the slower *Beth*). But it make no real difference whether we consider *AlephBeth* as one object or two, and so the intuition that *Aleph* reaches the ground before *Beth* must be mistaken.<sup>6</sup>

I will try to follow in Galileo's footsteps and argue that the intuition that *Alpha* might succeed while *Beta* fails is similarly mistaken. Consider *AlphaBeta*, a machine that works as follows. There are two trays on its left—one above the other—and a single tray on its right. *AlphaBeta* moves marbles from the lower left hand tray to the right hand tray. Initially there are infinitely many marbles in the upper left tray, a single marble in the lower left tray, and no marble on the right. Whenever *AlphaBeta* moves a marble, a hole opens up at the end of the upper tray and a single marble drops into the lower tray. *AlphaBeta* moves the marbles in the same pattern as *Alpha* and *Beta*, and it has the same task: to bring it about that all the marbles it is working with are on the right when it stops. It seems to me that (again, ignoring far-fetched possibilities and interruptions) the following claims are true:

- (1') Necessarily, *AlphaBeta* accomplishes its task if and only if *Alpha* does.
- (1'') Necessarily, *AlphaBeta* accomplishes its task if and only if *Beta* does.

One way to think about *AlphaBeta* is this: it moves infinitely many marbles from left to right in exactly the same pattern as *Alpha*. It makes no real difference whether the marbles that are waiting to be moved on the left will have dropped a bit just before the scoop picks them up. Another way to think about *AlphaBeta* is this: it moves a marble from the lower left tray to the right and while it rests, a marble shows up in the lower left tray, and this happens infinitely many times in the same pattern as with *Beta*. It makes no real difference whether the marble that shows up after each move is taken from the right hand tray or from some other place. But if both ways of thinking are legitimate then (1') and (1'') are true, and since they jointly entail (1), we now have an intuitive justification for the first premise.

In response, one could point out that there are possible laws that could guarantee that one of these machines fails without guaranteeing that the other does. For example, it could be a law that although things can move at any finite

<sup>6</sup> The idea of thinking about Galileo's thought experiment along these lines is from (Gendler 1998).



speed horizontally, they cannot move faster than the speed of light vertically; in a world governed by such a law *AlphaBeta* fails but *Alpha* might still succeed. Alternatively, it could be a law that after a single object oscillates between two locations infinitely many times, it goes out of existence; in a world with such a law, *Beta* fails but *AlphaBeta* might still succeed. But I assume that worlds governed by such *recherché* laws are beyond the scope of worlds we consider in making ordinary judgments of necessity. Note that Galileo's argument is subject to similar objections: it could be a law that whenever two objects are tied together they fuse into an extended simple, or go out of existence. All we can say in defense of the Galilean thought experiment is that such laws are far-fetched enough to be properly ignored.

I concede that there is a fairly natural law that distinguishes between *AlphaBeta* and *Beta*: the law of the continuity of motion. Let's assume that the law holds in an all nearby worlds. It follows that in nearby worlds, at the moment *Beta* and *Gamma* stop, the marble goes out of existence. (For if it existed, it would presumably have to be at a single location *l*. And at times arbitrarily close to the time when *Beta* and *Gamma* stopped, the marble had been at some fixed distance from *l*, which violates the continuity of motion.) On the other hand, there seems to be nothing that forces any of the marbles *AlphaBeta* moved from the left to the right to go out of existence. So, one might argue, in some nearby worlds *AlphaBeta* succeeds even though *Beta* fails, and so (1'') is false. But the last step in this reasoning is fallacious: even if we grant that when *AlphaBeta* stops the marbles are all on the right, it does not follow that *AlphaBeta* accomplished its task. There is a logical gap between the claim that *AlphaBeta* moved each marble from left to right and the claim that it moved the marbles from left to right (just as there is a logical gap between the claim that someone counted each guest and the claim that he counted the guests). Given the intuitive plausibility of (1''), I suggest that in nearby worlds *AlphaBeta* fails even though when it stops, the marbles are all on the right. So, accepting the claim that motion is necessarily continuous does not undermine the argument.

Before moving on, I'd like to restate Black's argument in terms of objective chances; this brings out the role of considerations about laws. Let  $P_x$  be a proposition describing the process machine *x* goes through and let  $S_x$  be a proposition describing the state the marbles must be in when *x* stops if *x* accomplishes its task. If a state is a result of a process, then the chances of the state holding must be higher than it not holding, given the process.<sup>7</sup> This means that *Alpha* cannot accomplish its task unless:

$$(5) \quad Pr(S_{Alpha} | P_{Alpha}) > Pr(\neg S_{Alpha} | P_{Alpha}).$$

<sup>7</sup> I am assuming here only a necessary (but not sufficient) condition for a state holding as a result of a process: if the state *S* holds as a result of the process *P*, then the objective chance of *S* holding after *P* is higher than the objective chance of *S* not holding after *P*.

I am using  $Pr$  for objective chances set by the laws. Worlds where the laws distinguish between *Alpha* and *Beta* are remote. So, in a world relevant for assessing ordinary necessity, (5) and (6) have the same truth-value:

$$(6) \quad Pr(S_{Beta} | P_{Beta}) > Pr(\neg S_{Beta} | P_{Beta}).$$

Now, by the logic of probability, (6) is equivalent to (7):

$$(7) \quad \frac{Pr(S_{Beta} \& \neg S_{Gamma} | P_{Beta}) + Pr(S_{Beta} \& S_{Gamma} | P_{Beta})}{Pr(\neg S_{Beta} \& \neg S_{Gamma} | P_{Beta}) + Pr(\neg S_{Beta} \& S_{Gamma} | P_{Beta})} >$$

Since *Beta* and *Gamma* work in tandem they go through the very same process. Accordingly,  $Pr(S_{Beta} \& \neg S_{Gamma} | P_{Beta})$  is the objective chance the marble ends up on the right and not on the left, given that these machines go through their moves. Similarly,  $Pr(\neg S_{Beta} \& S_{Gamma} | P_{Beta})$  is the objective chance that the marble ends up on the left, not on the right, given that these machines go through their moves. In worlds governed by normal laws these are the same. So, in those worlds (7) and (8) have the same truth-value:

$$(8) \quad Pr(S_{Beta} \& S_{Gamma} | P_{Beta}) > Pr(\neg S_{Beta} \& \neg S_{Gamma} | P_{Beta})$$

But (8) is false in worlds governed by normal laws: given that *Beta* and *Gamma* go through their moves, the objective chance that the marble ends up in *both* trays is certainly not higher than the objective chance that it ends up in *neither*. So, in the non-too-distant worlds relevant for assessing ordinary necessity (5) is also false, which means that *Alpha* does not accomplish its task.

#### 4 SUPER-TASKS AND ULTRA-TASKS

The strength of Black's argument has not been widely appreciated. This is, to a large extent, his own fault: he repeatedly misstated its conclusion. He says that the argument shows that it is logically impossible to perform an infinite series of tasks. He even says that "the notion of an infinite series of act is self-contradictory" (Black 1951, 101). The argument, as I presented it, shows no such thing. We are given no reason to think that *Alpha* cannot move *each marble* from left to right—the conclusion is merely that it cannot move *the marbles* from left to right. Suppose *Alpha* moves each of the marbles and after it is done the marbles are all on the right. This could happen. But if it does, the outcome does not come about as a result of the infinite series of tasks *Alpha* performs. Perhaps it happens as a result of some interference, or as a result of some other thing *Alpha* does. Or perhaps it happens not as a result of anything in particular.

The point can be clarified by distinguishing between two notions of an infinitary task. One is that of a *super-task* – a series of tasks of type  $\omega$  performed in a finite amount of time. The other is that of an *ultra-task* – a single task performed by performing a super-task.<sup>8</sup> Each individual move performed by *Alpha* is a task. If *Alpha* performs each of its moves it performs a super-task. What we earlier called the task of *Alpha* is an ultra-task: moving the marbles from left to right by performing this super-task. What Black's argument shows is that *Alpha* cannot perform this ultra-task.<sup>9</sup> And since there is nothing special about *Alpha*, a reasonable conjecture is that the same holds for ultra-tasks *tout court*: without there being some very odd laws, they cannot be performed at all.

The literature that followed Black's paper tended to focus on the question whether super-tasks are possible. The consensus appears to be that they are—and I share this view. But can one perform some other task by performing a super-task in the sense in which one can cross the street by making a series of steps or one can draw a picture by connecting a series of points? Considerations inspired by Black suggest a negative answer to this latter question. Super-tasks are *possible* but they are *inert*—when you perform a super-task you cannot thereby perform something over and above the individual tasks included in the super-task.

Hercules's second labor was to kill the Lernean Hydra, a nine-headed monster. The Hydra was a tough opponent: whenever Hercules cut off one of its heads, two new heads grew in its place. Killing such a creature by exponentially accelerating decapitation is an ultra-task, and as such, it is impossible to perform.<sup>10</sup> This is not to say that the supertask of cutting off infinitely many heads cannot *in principle* be performed, or that it is impossible for the Hydra to end up dead after this super-task is performed. Remarkable though it is, all this can happen: after the number of the Hydra's heads grows steadily beyond any finite limit, the Hydra suddenly finds itself headless. But this would not be a killing of the Hydra: the beast would not die as a result of what Hercules did. Setting aside the possibility of strange laws, some intervening force or miracle was also needed. Or perhaps its death was not the result of any process whatsoever.

<sup>8</sup> The notion of a *super-task* is due to (Thomson 1954). (Benacerraf 1962) introduced the notion of a *super-duper-task*: a series of tasks of type  $\omega + 1$ . Peter Clark and Stephen Read (1984) suggested the notion of a *hyper-task*: series of uncountably many tasks. While I am almost out of adjectives, an *ultra-task* is fundamentally different from all of these. It is not a series of tasks, just one task—constituted by infinitely many.

<sup>9</sup> I don't think Black saw clearly the distinction between super-tasks and ultra-tasks: this is the fundamental confusion in his paper.

<sup>10</sup> According to Apollodorus's account, Hercules manages to kill the Hydra with the help of his trusted nephew Iolaus. Every time Hercules cut off one of the heads, Iolaus held a torch to the stump preventing the growth of the new heads. So, Hercules killed the Hydra by performing a finite series of tasks.

Catching up with the tortoise by traversing infinitely many distinct intervals would also be an ultra-task—something that cannot be done. Of course, Achilles can catch up with the tortoise and—contra Black—he can also traverse infinitely many distinct intervals in a finite time. What Achilles cannot do is perform the former task by performing the latter. The “mathematical” resolution of Zeno’s paradox is wrong, even though both common sense and mathematics are vindicated. If you want an answer to the question “In virtue of what does Achilles catch up with the tortoise? (and it is by no means clear that you should want an answer to such a question) you need to appeal to something other than a super-task he performs. You can, for example, truthfully say that Achilles caught up with the tortoise by running faster than the animal, or by moving his legs one after the other, or by traversing a distance of some specific length.

## 5 TELICITY

I have argued that ultra-tasks are impossible (although not logically impossible) to perform. We can gain a better understanding of why this is so by examining what tasks are.

Let’s start with some examples. Catching up with a tortoise, moving a marble from one tray to another, killing the Hydra—these are all tasks. One thing they all have in common is that they all happen *in* a time and are not going *for* a time. For example, we say that Achilles caught up with the tortoise *in* two minutes, but not that he did that *for* two minutes. By contrast, we say that Achilles ran *for* an hour, but not that he ran *in* an hour. This is the classic test for telicity: it shows that ‘caught up with a tortoise’ a *telic* verb phrase and ‘run’ an *atelic* one.

Tasks are events described by telic verb phrases. Such events include a result, or *telos*, whose obtaining marks the end of the task. The *telos* of catching up with a tortoise is being lined up with the tortoise, the *telos* of moving a marble from one tray to another is for that marble to be in the latter tray, the *telos* of killing the Hydra is for the Hydra to be dead, and so on. I suggest that tasks are compound events consisting of a process leading up to a result state. Telic verb phrases describing a single task<sup>11</sup> say of an object that it is involved in a certain process that led to its *telos*. To catch up with a tortoise is to be doing something in a way that leads to being lined up with the tortoise, to move a marble from one tray to another is to be moving the marble from one tray in a way that leads to its being in the other, to kill the Hydra is to be killing it in a way that leads to its death, and so on. The phrase “counting the guests” in its most natural reading

<sup>11</sup> A verb phrase can be used to describe a single event or a multitude of events. ‘John traveled in four countries’ can mean that there was an event of him traveling in four countries or that four countries are such that there was an event of him traveling in them.

describes a single task; its *telos* is having an answer to the question “How many guests are there?” and the counting process leads up to the state of having this answer.

There is a distinction due to (Vendler 1967) within the category of telic verb phrases between *accomplishments* and *achievements*. The former are said to describe events that are extended in time (like stealing a car or building a house), and the latter events that are near-instantaneous (like reaching the peak or finding a key). It is sometimes said that achievement verb phrases don't allow the progressive, but this is not a reliable criterion (e.g. ‘Jack was finding his key’ is indeed odd, but ‘Jill was reaching the peak is just fine.) I think the real difference between them is that in the case of accomplishments, the process that leads up to the *telos* is properly described with the *progressive from of the verb phrase*, while this is not so in the case of achievements. So if Mary crossed the street (accomplishment) then she was crossing it and this particular process led to her being across the street. But if Mary got across the street (achievement), then something was going on—perhaps she was crossing the street, perhaps she was being carried by someone else, perhaps she was being teleported, it does not matter—and this process, whatever it was, led to her being across the street.

If these ideas are on the right path, then it is part of the meaning of telic verb phrases that a process leads to its natural result.<sup>12</sup> Moving infinitely many marbles from the left hand tray to the right hand tray means moving them from the left hand tray in a way that leads to their being in the right hand tray. This is no doubt possible: one could pick up all the marbles at once from the left hand tray and place them in the right hand one. But moving each of the marbles individually would be a super-task and, I argued, such a super-task does not lead to the marbles being in the right hand tray. The marbles may each be moved from the left and they may all end up on the right, but the latter would not happen as a result of the former. It may happen as a result of something else or not as a result of anything at all.

To count the *F*'s, I suggested, is to come up with an answer to the question “How many *F*'s are there?” *as a result* of counting each. Now we can see that this suggestion is a consequence of a general semantic thesis about telic verb phrases and a specific proposal about the *telos* of counting. Counting the guests at a birthday party is a straightforward task; counting the primes is not. The latter is an ultra-task, and as such, impossible. You could, in principle, count each prime and at the end come to have the answer that there are infinitely many of them. But your counting would not lead to your having that answer. The “countable” multitudes cannot be counted after all.\*

<sup>12</sup> For a sketch of a semantic account along these lines see (Szabó 2004) and (Szabó 2008).

\* A version of this paper was presented at the *Semantics and Philosophy in Europe* conference in Paris, at the *Arché/CS.M.N Graduate Conference* in Oslo, at the University of Connecticut

## REFERENCES

- Benacerraf, P., 1962, 'Tasks, Super-tasks and the Modern Eleatics. *Journal of Philosophy* 59, 765–84.
- Black, M. 1951, Achilles and the Tortoise. *Analysis* 11, 91–101.
- Clark, P. and S. Read, 1984, Hypertasks. *Synthese* 61, 387 – 390.
- Gendler, T. S. 1998, Galileo and the Indispensability of Scientific Thought Experiments. *The British Journal for the Philosophy of Science* 49, 397 – 424.
- Szabó, Z. G. 2004, On the Progressive and the Perfective. *Nous* 38, 29 – 5.
- Szabó, Z. G. 2008, Things in Progress. *Philosophical Perspectives* 22, 499 – 525.
- Thomson, J. 1954, Tasks and Super-tasks. *Analysis* 15, 1 – 10.
- Vendler, Z. 1967, Verbs and Times. In *Linguistics in Philosophy*, Ithaca, Cornell University Press, 97 – 121.

---

at Storrs, and at the memorial conference honoring Imre Ruzsa in Budapest. I thank the participants at each of these occasions for a lively and constructive discussion. I also thank Cian Dorr, Tamar Szabó Gendler, Daniel Rothschild, Ted Sider, and Brian Weatherson for written comments and criticism.

## What Mathematicians Say Means: In Defense of Hermeneutic Fictionalism<sup>1</sup>

**Abstract.** Hermeneutic fictionalism about mathematics maintains that mathematics is not committed to the existence of abstract objects such as numbers. Mathematical sentences are true, but they should not be construed literally. Numbers are just fictions in terms of which we can conveniently describe things which exist. The paper defends Stephen Yablo's hermeneutic fictionalism against an objection proposed by John Burgess and Gideon Rosen. The objection, directed against all forms of nominalism, goes as follows. Nominalism can take either a hermeneutic form and claim that mathematics, when rightly understood, is not committed to the existence of abstract objects, or a revolutionary form and claim that mathematics is to be understood literally but is false. The hermeneutic version is said to be untenable because there is no philosophically unbiased linguistic argument to show that mathematics should not be understood literally. Against this I argue that it is wrong to demand that hermeneutic fictionalism should be established solely on the basis of linguistic evidence. In addition, there are reasons to think that hermeneutic fictionalism cannot even be defeated by linguistic arguments alone.

Fictionalism is a general term for approaches which analyze a particular discourse or a particular idiom in terms of fictions. Take, for example, the sentence "The average star has 2.4 planets". Given the logical form of sentences involving definite descriptions, this sentence seems to assert that there is one and only one object which is the average star. But there is no such object, so the sentence is false. How come, then, that we find it true? The fictionalist says that in using this sentence we engage in a sort of game. We pretend that there is such an object and use this pretense to express a truth, namely, that if divide the number of planets with the number of stars we get 2.4.

<sup>1</sup>The research leading to this paper was supported by OTKA (National Foundation for Scientific Research), grant no. K 76865

Fictionalism can be pursued in a hermeneutic and in a revolutionary spirit.<sup>2</sup> Hermeneutic fictionalism seeks to uncover how the given discourse or idiom is in fact understood, i.e. to bring to the fore the meaning which has been there all along. The example just used is an instance of hermeneutic fictionalism. It does not tell us that we should stop believing in the existence of an average star, for we have never believed that. It tells us that instead of looking for a novel construal of the logical form of the sentence which would make it literally true, we should accept that it has the logical form it seems to have and it is not literally true.<sup>3</sup> Revolutionary fictionalism, in contrast, claims to reveal that what we took to be real is in fact a piece of fiction. It opens our eyes to the fact that we were wrong, and calls on us to change our commitments. Such is Field's attempt to counter Quine's and Putnam's indispensability argument, according to which we cannot but accept that the abstract objects of mathematics exist, because physics cannot do without them.<sup>4</sup> He attempts to show that physics can be pursued without numbers, so we do not have to put up with their existence.<sup>5</sup>

Stephen Yablo advocates hermeneutic fictionalism with respect to mathematics, and his theory has many attractions. It is nominalistic, so it can avoid the epistemological problem raised by Benacerraf. (A note of clarification: by 'nominalism' I mean the rejection of abstract objects and not the rejection of universals; nominalism so conceived is compatible with *in re* realism about universals.) In addition, it promises to explain why mathematics is necessary, how we can know it a priori, why we feel that mathematics is absolute in the sense that there cannot be an alternative arithmetic or set theory, why mathematics can be applied to the physical world, and many other things, including certain features of mathematical language. I will not elaborate on these, I will simply assume that it can deliver what it promises. In this paper I attempt to defend hermeneutic fictionalism against an objection first formulated by John Burgess, which he repeated several times, sometimes together with Gideon Rosen. I will start by a brief sketch of the account, which certainly will not do justice to its full complexity. Then I respond to the objection in two steps. Burgess and Rosen claim that the fate of hermeneutic fictionalism should be decided solely on the basis of empirical linguistic evidence. I argue first that the supportive evidence may come from philosophical considerations as well. Then I suggest, somewhat tentatively, that linguistic evidence alone might not even be sufficient for refutation.

<sup>2</sup> The hermeneutic-revolutionary distinction was introduced in (Burgess 2008a) and is first applied to fictionalism in (Stanley 2001).

<sup>3</sup> For a criticism of the fictionalist analysis of 'the average' example see (Stanley 2001, 54-58). For a response see (Yablo 2001, 93-96).

<sup>4</sup> (Quine 1980a, 1980b, 1981a, 1981b), (Putnam 1979a, 1979b).

<sup>5</sup> (Field 1980).



So let me start with Yablo. Quine has taught us that ontological commitment is marked by quantification. The entities whose existence we are committed to are the ones which we quantify over. Mathematics abounds with theorems which quantify over numbers, e.g. 'Any two numbers have a product'. It seems then that the truth of mathematical theorems implies that numbers exist. Yablo claims that quantifying over numbers incurs no such commitment just as by asserting that 'The average star has 2.4 planets', we do not incur commitment to the existence of the average star. But how can we quantify over numbers and yet abstain from ontological commitment?

Here is how. Number words have a use which is ontologically innocent, namely when they occur as devices of numerical quantification, like in 'There are twelve apostles'. Here the number word can be resolved into the standard devices of first order predicate logic with identity.<sup>6</sup> Starting from this innocent use we can get to quantification over numbers which is just as innocent by adopting a rule, which licenses the expression of the content of sentences involving numerical quantification in terms of quantification over numbers. Stated in a preliminary form, the rule says: if there are  $n$   $F$ 's, imagine there is a thing  $n$  which is identical with the number of  $F$ 's. Using  $*S*$  as notation to be read 'imagine/suppose that  $S$ ', the rule can be written as follows:

$(N_{\text{preliminary}})$  if  $\exists_n x (Fx)$ , then  $*\text{there is a thing } n (n = \text{the number of } F\text{'s})*^7$

$F$  is a predicate applicable to ordinary objects, and in the antecedent, we have a simple numerical quantification that does not assume the existence of numbers as objects. In the consequent we have quantification over numbers, but the quantification is ontologically innocent, since it occurs in the scope of the 'imagine that' operator. When we merely imagine that something exists, we are not committed to its existence. What the rule says is not that whenever a specifiable real world condition obtains, there exists a given number; it says that whenever a certain real world condition obtains we are allowed to engage in a game of make-belief and pretend that a given number exists.

This rule, however, will not quite do, because it does not allow us to assign numbers to numbers, like when we say 'The number of even primes equals 1'. 'Even' and 'prime' are predicates applicable to numbers, not to ordinary objects, so they cannot occur in the antecedent of the rule. We need to liberalize the rule and allow such predicates in the antecedent. But if we deny that numbers exist, we must also deny that the properties even and prime are instantiated. However, if we may imagine that numbers exist, we may also imagine that these properties are instantiated. This gives us a clue as to how the rule should be amended:

<sup>6</sup> There are  $n$   $F$ 's can be defined recursively as follows:  $\exists_n x Fx =_{\text{df}} \forall x (Fx \supset x \neq x)$ , and  $\exists_{n+1} x Fx =_{\text{df}} \exists y (Fy \ \& \ \exists_n x (Fx \ \& \ x \neq y))$ .

<sup>7</sup> The following account is based primarily on (Yablo 2002).

(N) if  $*\exists_n x (Fx)*$ , then  $*\text{there is a thing } n \text{ (} n = \text{the number of } F\text{'s)}*$

This rule says that if you imagine that there are  $n$   $F$ s, where  $F$  may be a property of ordinary objects or numbers, you may also imagine that there is an object which is the number of  $F$ s. This rule includes the preliminary one as a special case: if the antecedent of  $(N_{\text{preliminary}})$  is satisfied, i.e. if there are indeed a certain number of ordinary objects which are  $F$ , you are certainly entitled to imagine that.<sup>8</sup>

But why is it worth pretending that numbers exist? Because of the expressive power the quantificational idiom brings. Without this idiom, it would not be possible, for example, to formulate the laws of physics. Instead of Newton's second law, we could only formulate a huge conjunction with conjuncts of the form 'if a force  $F$  is exerted on a body with the mass  $M$ , it produces acceleration  $A$ '. But we would need an infinite number of conjuncts. Worse, since the magnitudes in question can take real numbers as values, the number of conjuncts should have to be uncountably infinite. If we are allowed to quantify over numbers, we can simply say, 'For all real numbers  $F$ ,  $M$  and  $A$ , if  $F$  = the force acting on a body with the mass =  $M$ , and  $A$  = the acceleration produced, then  $F = M \times A$ '.

It is exactly because of the expressive power of quantification over numbers that Quine believes that mathematical objects are indispensable for physics. Whereas Field accepts that the quantificational idiom yields ontological commitment, and tries to show that we can achieve the same expressive power without quantifying over numbers, Yablo maintains that we may quantify over numbers and yet avoid commitment. We simply pretend that there are mathematical entities. He points out that the use of fictions for purposes of representation is very common. For instance, you may describe a certain bodily feel of nervousness by saying 'There are butterflies in my stomach'. Of course, you do not believe that there are. But if there were, you think that would feel in this way. So you call us to imagine a fictitious state of affairs in order to describe a state of affairs which is real. Indeed, this is the way in which metaphors usually work. Metaphors, read literally, are typically false, but they call us to imagine something. If the call is accepted, the features of what is imagined point us to certain features of reality. One may describe the location of the city of Crotona saying 'It is on the arch of the Italian boot'.<sup>9</sup> Italy is not a boot, but if you are willing to pretend that it is, the sentence tells us where the city is to be found. It is because mathematics shares this feature of figurative speech that Yablo prefers to call his approach 'figuralism'.

<sup>8</sup> Once (N) is in place, we can have infinitely many numbers even if there are only finitely many ordinary objects. 0 is the number of things not identical to themselves,  $n$  is the number of numbers smaller than  $n$ .

<sup>9</sup> The example from (Walton 1993) 40-41, whose work is a major source of inspiration for fictionalism.

We have seen that real contents of sentences of applied mathematics are states of affairs which include nothing mathematical. But what about pure mathematics? What is, for instance ‘ $3 + 5 = 8$ ’ about if not about numbers? Yablo shows how sentences of pure mathematics can be recast in the ontologically innocent idiom of numerical quantification. The basic idea is to use rule (N) backwards. What the previous sentence really says is something like this: ‘If there are exactly three *F*’s and there are exactly five *G*’s, and no *F* is a *G*, then there are exactly eight objects which are *F*’s or *G*’s’. This is a logical truth. Yablo goes on to show how to reconstruct all sentences of arithmetic, including the ones which quantify over numbers, as logical truths, and he does the same for set theory. You can already see how Yablo can explain why mathematics is necessary and how it can be known a priori.

This should suffice to give us a flavor of Yablo’s approach. Let us now see why Burgess and Rosen believe that an account along these lines is untenable. The objection is not directed specifically against fictionalism but against nominalism in general. The nominalist denies the existence of abstract objects, so he does not accept that the mathematical sentences apparently asserting the existence of such objects are literally true. At this point, he has two options. To admit that these sentences are true and deny that they are understood literally, or to admit that they are understood literally and deny that they are true. The former is the hermeneutic, the latter is the revolutionary position. Burgess and Rosen argue that both are untenable. The hermeneutic position fails because it is not supported by *scientific* evidence. The revolutionary position fails because there are no sound scientific reasons to challenge the truth of mathematics or to replace current mathematics with a nominalistic alternative such as Field’s or Chihara’s. I emphasize “scientific”, because Burgess and Rosen are of the conviction that purely philosophical considerations can never take precedence over scientific reasoning. For example, epistemological worries about how we can acquire knowledge of the abstract entities of mathematics are not sufficient to discredit mathematicians’ claims to knowledge, and *a fortiori*, the truths of mathematics.<sup>10</sup> I grant this.

Nonetheless—and now I am starting with the response—when it comes to arguing against the hermeneutic approach, the point that purely philosophical considerations cannot trump scientific ones is replaced by something stronger, namely that philosophical considerations are simply irrelevant and carry no weight at all. They write “no nominalists favoring such a reconstrual have ever published their suggestions in a linguistics journal with evidence such as a linguist without ulterior ontological motives might accept”.<sup>11</sup> At another place Burgess briefly responds to those criticisms which allege that nominalists can have a third alternative in addition to hermeneutics and revolution.

<sup>10</sup> (Burgess and Rosen 2005, 520-523.)

<sup>11</sup> (Burgess and Rosen 2005, 525.)

[I]t is sometimes said that a nominalist interpretation represents “the best way to make sense of” what mathematicians say. I see in this formulation not a third alternative, but simply an equivocation, between “the empirical hypothesis about what mathematicians mean that best agrees with the evidence” (hermeneutic) and “the construction that can be put on mathematicians’ words that would best reconcile them with certain philosophical principles or prejudices” (revolutionary).<sup>12</sup>

What these remarks indicate is that the evidence for a nominalist interpretation of mathematics, such as Yablo’s, should be purely empirical and should not rely on philosophical considerations. This is actually how Burgess and Rosen proceed when they take up Yablo’s position.<sup>13</sup> They systematically ignore the philosophical benefits Yablo’s account may bring, and focus on the evidence from linguistic behavior. E.g. Yablo claims that the ease with which we pass from ontological innocent number talk to the quantificational formula, that we do not demand a proof existence, suggests that the latter idiom does not carry ontological commitment either. Or: if the Oracle mentioned in Burgess’ and Rosen’s book,<sup>14</sup> who knows exactly what exists, would proclaim that only concrete objects exist, mathematicians would not renounce their existence claims. I do not want to discuss Yablo’s linguistic arguments and Burgess’ and Rosen’s rejoinders. Suffice it to say that I do not find the rejoinders convincing, and I will later argue that a knockdown linguistic counterargument might not be that easy to formulate.

What I contend is that in assessing the case for hermeneutic fictionalism, it is wrong to disregard philosophical considerations.<sup>15</sup> I do not base this on the intrinsic importance of philosophy but on two facts about interpretation. First fact: interpretation—be it the interpretation of a text, of the behavior of a person, of a set social practices—is aimed at making sense, i.e. showing how the various parts hang together, how they cohere. The pursuit of coherence is checked against the empirical facts. Here is an example. Before the elections, a politician promises not to raise taxes, he comes to power, then raises them. There are several ways this may make sense. One: he believed he would not

<sup>12</sup> (Burgess 2008b, 51.)

<sup>13</sup> (Burgess and Rosen 2005, 528-534.)

<sup>14</sup> (Burgess and Rosen 1997, 3.)

<sup>15</sup> If I succeed, I shall have also disposed of Mark Balaguer’s objection. In Balaguer’s taxonomy there is no room for hermeneutic fictionalism. He defines fictionalism as the view that mathematical sentences should be taken at face value and are false. Yablo believes that mathematical sentences are true, so he is what Balaguer calls a paraphrase nominalist. Paraphrase nominalism is wrong because the empirical evidence suggests that mathematicians understand mathematical sentences literally and not according to the nominalist paraphrase. To me, this sounds like the same complaint as the one raised by Burgess and Rosen. (Balaguer 2008), (Balaguer 2009, 152, 158).

have to raise taxes and later found, to his dismay, that he was mistaken. Two: he knew all too well that he could not avoid raising taxes and calculated that the loss of credibility would be acceptable price for the increase of popularity the false promise would bring. Three: something in between; he was not certain, but he hoped he would not have to and took a calculated risk. Which is right? Empirical evidence decides. We have to find out what information he had about the state of the economy, how well he understood the information he had, what his advisors said, how often he kept his earlier promises, etc. And there are also several ways the story does not make sense (or at least does not make sense without further assumptions). One: he believed he would not have to raise taxes, and indeed he did not have to, still he raised them just for the fun of it. Two: he made a sincere promise and intended to keep it, just did not realize the legislation he passed was about tax raises. So an interpretation can fail in two ways: by conflicting with the empirical evidence and by violating the demand for coherence.

Second fact: judging whether or how much certain patterns are coherent draws heavily on the interpreter's own beliefs. This element of subjectivity is ineliminable, because there is no universal manual for identifying coherent patterns. The closest we have to such a manual is logic, but in matters of interpretation, logic might not have the last word. An interpretation which involves the attribution of inconsistency, might, on the whole, be better than one which involves the attribution a very far-fetched idea which happens to restore consistency. And to tell whether an idea is indeed far-fetched one has to rely on his own beliefs. Let me illustrate the same fact with the earlier example. Suppose you are thinking black and white. Then you will think that our politician either made a sincere promise but was unlucky, or he lied, and there are no other options. If you do think that, then, of course, you are a lousy interpreter. A good understanding of the field, human psychology and politics in this case, is necessary for a good interpretation. So the element of subjectivity does not imply arbitrariness.

How does this all bear on hermeneutic fictionalism? A philosopher, whose purpose is to interpret mathematics as a cognitive enterprise, wants to find out how various things in and around mathematics hang together. In deciding whether certain ideas cohere, he cannot but rely on what he believes. Suppose he believes that knowledge presupposes some kind of causal access. In that case, he would find it difficult to conceive how the Platonist account of mathematics, according to which mathematics provides literally true descriptions of abstract objects, which are not located in space-time and which are causally inert, may rationally cohere with the fact we do have mathematical knowledge. Or he may wonder how mathematics, alleged to describe causally inert objects, can benefit physics, which provides causal explanations.

If this is right, and the philosopher's interpretation of mathematics is a genuine interpretative enterprise, it cannot make do without reliance on the philosopher's

own convictions. So Burgess and Rosen are wrong when they demand that the interpretation of mathematics is to be based purely on empirical evidence, and should be free of philosophical considerations. Interpretation is never based purely on empirical evidence. It is in the business of uncovering coherence, rational connections between parts—and whether the parts are indeed rationally connected, is not something that can be empirically determined. The objection rests on a misunderstanding of what interpretation involves.

I want to emphasize that the above view of interpretation does not mean that philosophers are entitled to read into mathematics whatever philosophical views they happen to have. In order to see that, it is worth taking a look at how the empirical evidence and the interpreter's convictions interact in the course of interpretation. Suppose a historian is writing a book on Kepler. The dates when Kepler's books were published can be determined empirically. Once again, empirical evidence shows that the astronomical theory of *Harmonice Mundi*, which includes what we now call Kepler's laws, is superior to the astronomical theory in his first book, *Mysterium Cosmographicum*. But it is not empirical evidence which says that it was extremely odd of Kepler to republish his first book two years after the publication *Harmonice Mundi*. This judgment draws on the historian's own understanding that science aims primarily at empirically accurate theories. Given this understanding, the publication of an empirically inferior theory just does not make sense. The historian needs to find a coherent pattern which Kepler's actions fit. He may, for example suggest, that Kepler does not share the current view that empirical accuracy has exclusive importance. Kepler was a Platonist and held that that the world should exhibit an impressive mathematical order. Now *Mysterium Cosmographicum* is superior to *Harmonice Mundi* in terms of mathematical order. (Its leading idea is that orbits are circular and their distances are regulated by the five platonic solids: a platonic solid circumscribed around the orbit a planet closer to the Sun is inscribed in the orbit of the planet farther from the Sun.) Now, the historian who proceeds like this does not simply impute his own beliefs to Kepler, since he admits that Kepler's vision of science is different from his own. But he does not put his own beliefs aside either. After all, it is in terms of a belief he shares with Kepler that he makes sense of Kepler's actions, namely that it is right to publish what one believes to be good science. It is in the light of this conviction that Kepler's actions turn out in a way rational. So the way to conceive the role of the interpreter's own convictions is this. The interpreter's convictions provide ways in which what is interpreted can be construed as exhibiting coherence. The role of empirical evidence is to determine which ones of these coherent patterns are, in fact, exhibited.

I have been arguing so far that Burgess and Rosen cannot rule that evidence from philosophical considerations is inadmissible. This may strengthen the case for hermeneutic fictionalism. Now I want to go further and suggest that it is not entirely clear that hermeneutic fictionalism can be refuted at all solely

by non-philosophical considerations. Suppose we consider only arguments from the mathematicians' linguistic behavior and in the interpretation of what mathematicians say and write we consciously abstain from relying on philosophical considerations. I will consider two scenarios in which the result of such non-philosophical arguments is apparently unfavorable to hermeneutic fictionalism and claim that these scenarios do not suffice to refute hermeneutic fictionalism.

The first scenario is that we find that mathematicians do not believe that mathematical objects are fictions because they do not have beliefs about their ontological status. For instance, an empirical survey shows that the overwhelming majority of mathematicians say that they have not thought much about this question, they are not particularly interested in it, or claim to be ignorant about it, or are ready to adopt any position recommended to them; and the minority which displays interest consist of two groups. Members of the one have views which are vague, ambiguous, inconsistent or otherwise unsatisfactory. Members of the other minority group are very sophisticated but cannot agree among themselves. If this were the case, the hermeneutic fictionalist would have to choose carefully the way in which he formulates his position. In particular, he should make it very clear that he is not offering a psychological description of what mathematicians think. He should possibly avoid talking about mathematicians' beliefs, or explain that what he calls beliefs are the views which make best sense of what mathematicians do rather than the dispositional mental states they have. Or he should prefer to talk about mathematics and mathematical practice rather than of mathematicians.

This would not be an ad hoc maneuver. Interpretations often involve elements which are not meant to be psychologically faithful.<sup>16</sup> As a first example, take some current interpretations of Descartes which allege that ideas are to be understood as intentional contents. Viewed as a psychological statement, this would involve some distortion, because Descartes did not possess the concept of intentional content. Today's concept of intentional content is informed by the tradition of Brentano, Husserl, Frege and Chisholm, which emerged only much later. Instead, advocates of this interpretation should be viewed as claiming that understanding Descartes's concept as intentional content is consistent with what Descartes actually says, and sheds light on how various elements of Descartes's thought hang together. For a more dramatic example, take the interpretation of potlatch as a means of maintaining hierarchical relations between clans or villages. Surely, when the Indians of the Pacific Northwest gather to give away

<sup>16</sup> For the purposes of discussion, I assume two things. First, that facts about beliefs are as "hard" as any physical fact. Second, that people are not mistaken about their beliefs. Giving up either assumptions would give me more room to maneuver but would also invite several objections.

and often destroy vast amount of goods, they do not think of this as a way of reinforcing their social status. It is not just that they do not possess the concepts of social science. Even if they did, the social scientists' explanation, which is thoroughly secular, would not occur to them, because in their eyes, potlatch has a religious character.

It is important to see that interpretations which fail in terms of psychological faithfulness may be fully legitimate as interpretations—they are not abnormal or deviant. Interpretation is in the business of making sense, displaying how things rationally cohere. Now sometimes we cannot capture coherence in terms of the actual beliefs of the people we interpret. Descartes's concept of idea is not sufficiently clear to make the coherence of his thought transparent. The people practicing potlatch explain this custom in terms of following the law. But we believe that laws must serve some purpose, so we need a rationale, and the people do not provide one. If an interpreter finds that coherence cannot be captured by psychologically faithful descriptions, he forgoes psychological faithfulness. Similarly, if hermeneutic fictionalism succeeds in making sense of mathematics and its use in physics, it should not be faulted on grounds that it does not represent mathematicians' beliefs.

Let us move over to the second scenario. Here, the interpretation of the linguistic behavior of mathematicians—which relinquishes philosophical considerations—makes it clear that mathematicians reject fictionalism. Imagine, it turns out, they are all Platonists.

Notice that this would not automatically refute fictionalism. It might be the case that the fictionalists are right, and the mathematicians are wrong. This is Mark Balaguer's favored response to the Burgess-Rosen argument.<sup>17</sup> He claims that revolutionary fictionalism, which accepts that mathematical statements are understood literally and are false is tenable. It would be admissible to overrule mathematicians' judgments concerning the ontological status of mathematical entities, for two reasons. First, such a decision would be of little significance for mathematical practice. Second, mathematicians' professional expertise, which a philosopher cannot question, does not extend to the issues of ontological status.

However, hermeneutic fictionalism holds that mathematical statements are true, but are not understood literally, and it is hermeneutic fictionalism I wish to defend. There are two forms the defense can take. One is to reevaluate the mathematician's alleged commitment to Platonism. Suppose mathematicians explain why they take mathematical sentences literally true in the following way. "Look, we know how to tell metaphors from literal speech. We speak literally when we use the words as we ordinarily do. Now the word 'Sun' normally refers

<sup>17</sup> (Balaguer 2009, 153-157.); he believes though that there might also be a way to reject the hermeneutic-revolutionary distinction, 157-161.



to a hot ball of gas. When Romeo calls Juliet the Sun, he cannot be talking literally, since he cannot possibly believe that Juliet is a hot ball of gas. But as opposed to the word ‘Sun’, mathematical terms do not have an established use with which our use could be contrasted. So we are talking literally.” In response to this, the hermeneutic fictionalist may point out that certain expressions are inherently metaphorical in the sense that they do not have literal uses. Take the word ‘Vulcan’ introduced in Star Trek. If you call someone who always behaves in a cool, emotionally detached and highly logical fashion a Vulcan, you do not mean that he comes from a humanoid race which evolved on the planet Vulcan, since you know all too well that he does not. Or if you describe someone prone to emotional and illogical behavior as not being a Vulcan, you do not mean to assert that he does not from that race. And even if you call Captain Spock a Vulcan, you do not mean in all seriousness that there is an individual bearing this name who comes from the planet Vulcan. This example is meant to illustrate that when mathematicians confess to Platonism, that may be due to the fact that they misconstrue ‘literal’ or construe it in a way that differs from the hermeneutic fictionalist’s intention.<sup>18</sup>

But suppose no such maneuver is possible. Mathematicians happen to be very sophisticated in matters of linguistics, they do not misconstrue hermeneutic fictionalism, but they reject it in full knowledge of what it involves. That alone would still not be enough to refute hermeneutic fictionalism. When defending revolutionary fictionalism, Balaguer considers the idea that his revolutionism might not concern mathematics at all.<sup>19</sup> He envisages a version of Platonism which runs as follows. Mathematical facts are compounded of two sorts of facts: ontologically neutral facts about the correctness of mathematical sentences construed in fictionalist terms, and platonic facts to the effect that the abstract objects mathematical sentences seem to describe exist, which make it the case that the sentences which are correct in the fictionalist terms are actually true. It is only these platonic facts which on the fictionalist view do not obtain. Balaguer wonders if the platonic facts are mathematical facts at all. If not, the fictionalism he proposes would amount to a revolution in philosophy rather than mathematics. He admits that he does not know how to show that the alleged platonic facts are not mathematical in nature, and neither do I.

I believe, however, that the hermeneutic fictionalist can make a similar move and is in a position to argue for it. Suppose that if we take into account philosophical considerations and no others, fictionalism scores better than other alternatives. This should be granted for the sake of argument, since if fictionalism fails on philosophical grounds, it fails, and there is no point in trying to show that it can be maintained in the face of its rejection by mathematicians.

<sup>18</sup> For a more inclusive discussion see (Yablo 2000, 221-224.)

<sup>19</sup> (Balaguer 2009, 156.)

Then from the hermeneutic fictionalist's point of view, the situation looks as follows. Certain things mathematicians say, e.g. 'For every prime number there is a larger one', are true, even though not in a literal sense. Other things they say, e.g. 'Numbers are abstract objects and they do exist' are false in the literal sense. (If mathematicians did not intend these sentences in the literal sense, they would not be contradicting the fictionalist.) For sentences in the first group, they have arguments, which are virtually impossible to resist, and these arguments apply a small group of very special methods, such as deduction from axioms. Arguments for the sentences in the second group are not based on these special methods, and they can and should be resisted. Add to these certain behavioral or, if you wish, sociological facts. The professional training mathematicians receive prepares them to deal with the first group. The scholarly journals they publish in are devoted to the first group. One may gain recognition as a great mathematician only by establishing claims in the first group. Those who are exclusively concerned with the second group are typically not regarded as mathematicians, and the list may be continued. All in all, we find that the distinction between the two groups of sentences is not a local phenomenon but is manifested in many ways. Given the significance this distinction seems to have, an interpretation of mathematical practice has to account for it. And the easiest way to account for it is to say that sentences in the first group are the only ones that genuinely belong to mathematics. If this is right, then the mathematicians' uniform commitment to Platonism envisaged in this second scenario does not provide much of an argument against hermeneutic fictionalism, because this commitment falls outside territory of mathematics.

Let me summarize. I argued that Burgess and Rosen are wrong when they demand that hermeneutic fictionalism should be established purely by linguistic considerations. This argument was based on the nature of interpretation. I also raised doubts whether hermeneutic fictionalism can be defeated purely by linguistic considerations. I did that by considering two scenarios which might have seemed to support decisive linguistic objections. This latter argument was not meant to be conclusive. Perhaps one may develop a very well motivated account of fictional talk and use this to show that hermeneutic fictionalism is untenable.

## REFERENCES

- Balaguer, Mark, 2008, Fictionalism in the philosophy of mathematics. in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, URL = <<http://plato.stanford.edu/archives/fall2008/entries/fictionalism-mathematics/>>.
- Balaguer, Mark, 2009, Fictionalism, theft, and the story of mathematics. *Philosophia Mathematica* 17, 131-162.

- Burgess, John P., 2008a, Why I am not a nominalist. Reprinted in *Mathematics, Models, and Modality: Selected Philosophical Essays*, Cambridge: Cambridge University Press, 31-45.
- Burgess, John P., 2008b, Mathematics and bleak house. Reprinted in *Mathematics, Models, and Modality: Selected Philosophical Essays*, Cambridge: Cambridge University Press, 46-65.
- Burgess, John P. and Gideon Rosen, 1997, *A Subject With No Object*, New York, Oxford University Press.
- Burgess, John P. and Gideon Rosen, 2005, Nominalism reconsidered. In S. Shapiro (ed.), *The Oxford Handbook of Philosophy and Mathematics and Logic*. Oxford, Oxford University Press, 515- 535.
- Chihara, Charles, 2005, Nominalism. In *The Oxford Handbook of Philosophy and Mathematics and Logic*, Oxford, Oxford University Press, 483-514.
- Field, Harry, 1980, *Science Without Numbers*. Princeton (NJ), Princeton University Press.
- Putnam, Hilary, 1979a, What is mathematical truth? In *Mathematics, Matter and Method: Philosophical Papers* vol 1, 2<sup>nd</sup> ed., Cambridge, Cambridge University Press, 60-78.
- Putnam, Hilary, 1979b, Philosophy of logic. Reprinted in *Mathematics, Matter and Method: Philosophical Papers*, vol 1, 2<sup>nd</sup> ed., Cambridge, Cambridge University Press, 323-357.
- Quine, Willard Van Orman, 1980a, On what there is. Reprinted in *From a Logical Point of View*, 2nd ed., Cambridge (MA), Harvard University Press, 1-19.
- Quine, Willard Van Orman, 1980b, Two dogmas of empiricism. Reprinted in *From a Logical Point of View*, 2nd ed., Cambridge (MA), Harvard University Press, 20-46.
- Quine, Willard Van Orman, 1981a, Things and their place in theories. In *Theories and Things*. Cambridge (MA), Harvard University Press, pp. 1-23.
- Quine, Willard Van Orman, 1981b, Five milestones of empiricism. In *Theories and Things*, Cambridge (MA), Harvard University Press, 67-72.
- Stanley, Jason, 2001, Hermeneutic fictionalism. In P. A. French and H. K. Wettstein (eds.), *Midwest Studies in Philosophy 25: Figurative Language*, 36-71.
- Walton Kendall L., 1990, *Mimesis as Make-Believe: On the Foundation of the Representational Arts*. Cambridge (MA) / London, Harvard University Press.
- Walton Kendall L., 1993, Metaphor and prop oriented make-believe, *European Journal of Philosophy* 1, 39-57.
- Yablo, Stephen, 2000, Apriority & existence. In P. Boghossian and C. Peacocke (eds.) *New Essays on the A Priori*, Oxford: Oxford University Press, 197-228.
- Yablo, Stephen, 2001, Go figure: a path through fictionalism. In P. A. French and H. K. Wettstein (eds.), *Midwest Studies in Philosophy 25: Figurative Language*, 72-102.
- Yablo, Stephen, 2002, Abstract objects: a case study. *Nous* 36, supplementary volume 1, 220-240.
- Yablo, Stephen, 2005, The myth of the seven. In M. Kalderon (ed.), *Fictionalism in Metaphysics*, New York, Oxford University Press, pp. 88-115.
- Yablo, Stephen and Andre Gallois, 1998, Does ontology rest on a mistake?. *Proceedings of the Aristotelian Society*, supplementary volume 72, 229-262.

## On Logical Analysis of Relativity Theories

**Abstract.** The aim of this paper is to give an introduction to our axiomatic logical analysis of relativity theories.

### 1 INTRODUCTION

Our general aim is to build up relativity theories as theories in the sense of mathematical logic. So we axiomatize relativity theories within pure first-order logic (FOL) using simple, comprehensible and transparent basic assumptions (axioms). We strive to prove all the surprising predictions of relativity from a minimal number of convincing axioms. We eliminate tacit assumptions from relativity by replacing them with explicit axioms (in the spirit of the foundation of mathematics and Tarski's axiomatization of geometry). We also elaborate logical and conceptual analysis of our theories.

Logical axiomatization of physics, especially that of relativity theory, is not a new idea, among others, it goes back to such leading scientists as Hilbert, Reichenbach, Carnap, Gödel, and Tarski. Relativity theory was intimately connected to logic from the beginning, it was one of the central subjects of logical positivism. For a short survey on the broader literature, see, e.g., (Andréka & *al.* 2006). Our aims go beyond these approaches in that along with axiomatizing relativity theories we also analyze in detail their logical and conceptual structure and, in general, investigate them in various ways (using our logical framework as a starting point).

A novelty in our approach is that we try to keep the transition from special relativity to general relativity logically transparent and illuminating. We “derive” the axioms of general relativity from those of special relativity in two natural steps. First we extend our axiom system for special relativity with accelerated observers

(sec.7). Then we eliminate the distinguished status of inertial observers at the level of axioms (sec.8).

Some of the questions we study to clarify the logical structure of relativity theories are:

- What is believed and why?
- Which axioms are responsible for certain predictions?
- What happens if we discard some axioms?
- Can we change the axioms and at what price?

Our aims stated in the first paragraph reflect, partly, the fact that we axiomatize a physical theory. Namely, in physics the role of axioms (the role of statements that we assume without proofs) is more fundamental than in mathematics. Among others, this is why we aim to formulate simple, logically transparent and intuitively convincing axioms. Our goal is that on our approach, surprising or unusual predictions be theorems and not assumed as axioms. For example, the prediction “no faster than light motion ...” is a theorem on our approach and not an axiom, see Thm.5.1.

Getting rid of unnecessary axioms is especially important in a physical theory. When we check the applicability of a physical theory in a situation, we have to check whether the axioms of the theory hold or not. For this we often use empirical facts (outcomes of concrete experiments). However, these correspond to existentially quantified theorems<sup>1</sup> rather than to universally quantified statements—which the axioms usually are. Thus while we can easily disprove the axioms by referring to empirical facts, we can verify these axioms only to a certain degree. Some of the literature uses the term ‘empirical fact’ for universal generalization of an empirical fact elevated to the level of axioms, see, e.g., (Gömöri–Szabó 2010, §4), (Szabó 2009). We simply call these generalizations (empirical) axioms.

## 2 WHY RELATIVITY?

For one thing, Einstein’s theory of relativity not just had but still has a great impact on many areas of science. It has also greatly affected several areas in the philosophy of science. Relativity theory has an impact even on our every day life, e.g., via GPS technology (which cannot work without relativity theory). Any theory with such an impact is also interesting from the point of view of axiomatic foundations and logical analysis.

Since spacetime is a similar geometrical object as space, axiomatization of relativity theories (or spacetime theories in general) is a natural continuation of

<sup>1</sup>We do not want to assume every experimental fact as an axiom. We only want them to be consequences of our theories.

the works of Euclid, Hilbert, Tarski and many others axiomatizing the geometry of space.

### 3 WHY AXIOMATIC METHOD?

There are many examples showing the benefits of using axiomatic method. For example, if we decompose relativity theories into little parts (axioms), we can check what happens to our theory if we drop, weaken or replace an axiom or we can take any prediction, such as the twin paradox, and check which axiom is and which is not needed to derive it. This kind of reverse thinking helps to answer the why-type questions. For details on answering why-type questions by the methodology of the present work, see (Andréka & al. 2002, 12–13.), (Székely 2010a).

The success story of axiomatic method in the foundations of mathematics also suggests that it is worth applying this method in the foundations of spacetime theories (Friedman 2004a), (Friedman 2004b). Let us note here that Euclid's axiomatic-deductive approach to geometry also made a great impression on the young Einstein, see (Herschbach 2008).

Among others, logical analysis makes relativity theory modular: we can change some axioms, and our logical machinery ensures that we can continue working in the modified theory. This modularity might come handy, e.g., when we want to unify general relativity and quantum theory to a theory of quantum gravity. For further reasons why to apply the axiomatic method to spacetime theories, see, e.g., (Andréka & al. 2006), (Andréka & al. 2002), (Guts 1982), (Schutz 1973), (Suppes 1968).

### 4 WHY FIRST-ORDER LOGIC?

We aim to provide a logical foundation for spacetime theories similar to the rather successful foundations of mathematics, which, for good reasons, was performed strictly within FOL. One of these reasons is that FOL helps to avoid tacit assumptions. Another is that FOL has a complete inference system while second-order logic (or higher-order logic) cannot have one.

Still another reason for choosing FOL is that it can be viewed as a fragment of natural language with unambiguous syntax and semantics. Being a *fragment of natural language* is useful in our project because one of our aims is to make relativity theory accessible to a broad audience. *Unambiguous syntax and semantics* are important, because they make it possible for the reader to always know what is stated and what is not stated by the axioms. Therefore they can use the axioms without being familiar with all the tacit assumptions and rules of thumb of physics (which one usually learns via many, many years of practice).

For further reasons why to stay within FOL when dealing with axiomatic foundations, see, e.g., (Andréka & *al.* 2002, §Appendix: Why FOL?), (Ax 1978), (Székely 2009, §11), (Väänänen 2001), (Wolenski 2004).

## 5 SPECIAL RELATIVITY

Before we present our axiom system let us go back to Einstein's original (logically non-formalized) postulates. Einstein based his special theory of relativity on two postulates, the principle of relativity and the light principle: "The laws by which the states of physical systems undergo change are not affected, whether these changes of state be referred to the one or the other of two systems of coordinates in uniform translatory motion." and "Any ray of light moves in the 'stationary' system of co-ordinates with the determined velocity  $c$ , whether the ray be emitted by a stationary or by a moving body.", see (Einstein 1905/1952).

The logical formulation of Einstein's principle of relativity is not an easy task since it is difficult to capture axiomatically what "the laws of nature" are in general. Nevertheless, the principle of relativity can be captured by our FOL approach, see (Andréka & *al.* 2002), (Madarász 2002, §2.8.3).

Instead of formulating the two original principles, we formulate the following consequence of theirs: "the speed of light signals is the same in every direction everywhere according to every inertial observer" (and not just according to the 'stationary' observer). Here we will base our axiomatization on this consequence and call it light axiom. We will soon see that the light axiom can be regarded as the key assumption of special relativity.

Since we want to axiomatize special relativity, we have to fix some formal language in which we will write up our axioms. Let us see the basic concepts (the "vocabulary" of the FOL language) we will use. We would like to speak about motion. So we need a basic concept of things that can move. We will call these object *bodies*.<sup>2</sup> The light axiom requires a distinguished type of bodies called *photons* or *light signals*.<sup>3</sup> We will represent motion as the changing of spatial location in time. Thus we will use reference frames for coordinatizing events (meetings of bodies). Time and space will be marked by *quantities*. The structure of quantities will be an *ordered field* in place of the field of real numbers.<sup>4</sup> For

<sup>2</sup>By bodies we mean anything which can move, e.g., test-particles, reference frames, electromagnetic waves, etc.

<sup>3</sup>Here we use light signals and photons as synonyms because it is not important here whether we think of them as particles or electromagnetic waves. The only thing that matters here is that they are "things that can move." So they are bodies in the sense of our FOL language.

<sup>4</sup>Using ordered fields in place of the field of real numbers increases the flexibility of the theory and reduces the amount of mathematical presuppositions. For further motivation in

simplicity, we will associate special bodies to reference frames. These special bodies will be called “observers.” Observations will be formalized/represented by means of the *worldview relation*.

To formalize the ideas above, let us fix a natural number  $d \geq 2$  for the dimension of spacetime. To axiomatize theories of the  $d$ -dimensional spacetime, we will use the following two-sorted FOL language:

$$\{ B, \text{IOb}, \text{Ph}, Q, +, \cdot, W \},$$

where  $B$  (bodies) and  $Q$  (quantities) are the two sorts,<sup>5</sup> IOb (inertial observers) and Ph (light signals or photons) are one-place relation symbols of sort  $B$ ,  $+$  and  $\cdot$  are two-place function symbols of sort  $Q$ , and  $W$  (the worldview relation) is a  $2 + d$ -place relation symbol the first two arguments of which are of sort  $B$  and the rest are of sort  $Q$ .

Atomic formulas IOb( $k$ ) and Ph( $p$ ) are translated as “ $k$  is an inertial observer,” and “ $p$  is a photon,” respectively. To speak about coordinatization, we translate  $W(k, b, x_1, \dots, x_{d-1}, t)$  as “body  $k$  coordinatizes body  $b$  at space-time location  $\langle x_1, \dots, x_{d-1}, t \rangle$ ,” (i.e., at space location  $\langle x, \dots, x_{d-1} \rangle$  and at instant  $t$ ). Sometimes we use the more picturesque expressions *sees* or *observes* for *coordinatizes*. However, these cases of “seeing” and “observing” have nothing to do with visual seeing or observing; they only mean associating coordinate points to bodies.

The above, together with statements of the form  $x = y$  are the so-called *atomic formulas* of our FOL language, where  $x$  and  $y$  can be arbitrary variables of the same sort, or terms built up from variables of sort  $Q$  by using the two-place operations  $\cdot$  and  $+$ . The *formulas* are built up from these atomic formulas by using the logical connectives *not* ( $\neg$ ), *and* ( $\wedge$ ), *or* ( $\vee$ ), *implies* ( $\rightarrow$ ), *if-and-only-if* ( $\leftrightarrow$ ) and the quantifiers *exists* ( $\exists$ ) and *for all* ( $\forall$ ). For the precise definition of the syntax and semantics of FOL, see, e.g., (Chang–Keisler 1990, §1.3).

To meaningfully formulate the light axiom, we have to provide some algebraic structure for the quantities. Therefore, in our first axiom, we state some usual properties of addition  $+$  and multiplication  $\cdot$  true for real numbers.

**AxFd:** The quantity part  $\langle Q, +, \cdot \rangle$  is a Euclidean field, i.e.,

- $\langle Q, +, \cdot \rangle$  is a field in the sense of abstract algebra,
- the relation  $\leq$  defined by  $x \leq y \iff \exists z \ x + z^2 = y$  is a linear ordering on  $Q$ , and
- Positive elements have square roots:  $\forall x \ \exists y \ x = y^2 \vee -x = y^2$ .

this direction, see, e.g., (Ax 1978). Similar remarks apply to our other flexibility-oriented decisions, e.g., to treat the dimension of spacetime as a variable.

<sup>5</sup>That our theory is two-sorted means only that there are two types of basic objects (bodies and quantities) as opposed to, e.g., set theory where there is only one type of basic objects (sets).



The field-axioms (see, e.g., (Chang–Keisler 1990, 40–41.)) say that  $+$ ,  $\cdot$  are associative and commutative, they have neutral elements  $0$ ,  $1$  and inverses  $-$ ,  $/$  respectively, with the exception that  $0$  does not have an inverse with respect to  $\cdot$ , as well as  $\cdot$  is additive with respect to  $+$ . We will use  $0$ ,  $1$ ,  $-$ ,  $/$ ,  $\sqrt{\quad}$  as derived (i.e., defined) operation symbols.

AxFd is a “mathematical” axiom in spirit. However, it has physical (even empirical) relevance. Its physical relevance is that we can add and multiply the outcomes of our measurements and some basic rules apply to these operations. Physicists usually use all properties of the real numbers tacitly, without stating explicitly which property is assumed and why. The two properties of real numbers which are the most difficult to defend from an empirical point of view are the Archimedean property, see (Rosinger 2008), (Rosinger 2009, §3.1), and the supremum property,<sup>6</sup> see the remark after the introduction of axiom Cont on p.14.

Euclidean fields got their name after their role in Tarski’s FOL axiomatization of Euclidean geometry (Tarski 1959). By AxFd we can reason about the Euclidean structure of a coordinate system the usual way, we can introduce Euclidean distance, speak about straight lines, etc. In particular, we will use the following notation for  $\bar{x}, \bar{y} \in Q^n$  (i.e.,  $\bar{x}$  and  $\bar{y}$  are  $n$ -tuples over  $Q$ ) if  $n \geq 1$ :

$$|\bar{x}| \stackrel{d}{=} \sqrt{x_1^2 + \dots + x_n^2}, \quad \text{and} \quad \bar{x} - \bar{y} \stackrel{d}{=} \langle x_1 - y_1, \dots, x_n - y_n \rangle.$$

We will also use the following two notations:

$$\bar{x}_s \stackrel{d}{=} \langle x_1, \dots, x_{d-1} \rangle \quad \text{and} \quad x_t \stackrel{d}{=} x_d$$

for the *space component* and the *time component* of  $\bar{x} = \langle x_1, \dots, x_d \rangle \in Q^d$ , respectively.

Now let us see how the light axiom can be formalized in our FOL language.

**AxPh:** For any inertial observer, the speed of light is the same in every direction everywhere, and it is finite. Furthermore, it is possible to send out a light signal in any direction. Formally:

$$\forall m \exists c_m \forall \bar{x}\bar{y} \text{IOb}(m) \rightarrow (\exists p \text{Ph}(p) \wedge W(m, p, \bar{x}) \wedge W(m, p, \bar{y})) \leftrightarrow |\bar{y}_s - \bar{x}_s| = c_m \cdot |y_t - x_t|.$$

Axiom AxPh has an immediate physical meaning. This axiom is not only implied by the two original principles of relativity, but it is well supported by experiments, such as the Michelson-Morley experiment. Moreover, it has been continuously tested ever since then. Nowadays it is tested by GPS technology.

<sup>6</sup>The supremum property (i.e., every nonempty and bounded subset of the real numbers has a least upper bound) implies the Archimedean property. So if we want to get ourselves free from the Archimedean property, we have to leave this property, too.

Axiom AxPh says that “It is *possible* for a photon to move from  $\bar{x}$  to  $\bar{y}$  iff ...”. So, a notion of possibility plays a role here. In the present paper we work in an extensional framework, as is customary in geometry and in spacetime theory. However, it would be more natural to treat this “possibility phenomenon” in a modal logic framework, and this is more emphatically so for relativistic dynamics (Andréka & *al.* 2008). It would be interesting to explore the use of modal logic in our logical analysis of relativity theory. This investigation would be a nice unification of the works of Imre Ruzsa’s school on modal logic and the works of our Tarskian spirited school on axiomatic foundations of relativity theory. Robin Hirsch’s work can be considered as a first step along this road (Hirsch 2009).

Let us note that AxPh does not require that the speed of light be the same for every inertial observer or that it be nonzero. It requires only that the speed of light according to a fixed inertial observer be a quantity which does not depend on the direction or the location.

Why do we not require that the speed of light is nonzero? The main reason is that we are building our logical foundation of spacetime theories examining thoroughly each part of each axiom to see where and why we should assume them. Another (more technical) reason is that it will be more natural to include this assumption ( $c_m \neq 0$ ) in our auxiliary axiom AxSm on page 8.

Our next axiom connects the worldviews of different inertial observers by saying that all observers observe the same “external” reality (the same set of events). Intuitively, by the event occurring for  $m$  at  $\bar{x}$ , we mean the set of bodies  $m$  observes at  $\bar{x}$ . Formally:

$$\text{ev}_m(\bar{x}) \stackrel{d}{=} \{b : W(m, b, \bar{x})\}.$$

AxEv: All inertial observers coordinatize the same set of events:

$$\forall mk \text{ IOb}(m) \wedge \text{IOb}(k) \rightarrow \forall \bar{x} \exists \bar{y} \forall b W(m, b, \bar{x}) \leftrightarrow W(k, b, \bar{y}).$$

This axiom is very natural and tacitly assumed in the non-axiomatic approaches to special relativity, too.

Basically we are done. We have formalized the light axiom AxPh. We have introduced two supporting axioms (AxFd and AxEv) for the light axiom which are simple and natural; however, we cannot simply omit them without losing some of the meaning of AxPh. The field axiom enables us to speak about distances, time differences, speeds, etc. The event axiom ensures that different inertial observers see the same events.

In principle, we do not need more axioms for analyzing/axiomatizing special relativity, but let us introduce two more simplifying ones. We could leave them out without losing the essence of our theory, it is just that the formalizations of the theorems would become more complicated.

AxSf: Any inertial observer sees himself on the time axis:

$$\forall m \text{ IOb}(m) \rightarrow (\forall \bar{x} \ W(m, m, \bar{x}) \leftrightarrow x_1 = 0 \wedge x_2 = 0 \wedge x_3 = 0).$$

The role of AxSf is nothing more than making it easier to speak about the motion of reference frames via the motion of their time axes. Identifying the motion of reference frames with the motion of their time axes is a standard simplification in the literature. AxSf is a way to formally capture this simplifying identification.

Our last axiom is a symmetry axiom saying that all inertial observers use the same units of measurements.

AxSm: Any two inertial observers agree about the spatial distance between two events if these two events are simultaneous for both of them; furthermore, the speed of light is 1:

$$\begin{aligned} \forall mk \text{ IOb}(m) \wedge \text{IOb}(k) &\rightarrow \forall \bar{x}\bar{y}\bar{x}'\bar{y}' \ x_t = y_t \wedge x'_t = y'_t \wedge \\ \text{ev}_m(\bar{x}) = \text{ev}_k(\bar{x}') \wedge \text{ev}_m(\bar{y}) = \text{ev}_k(\bar{y}') &\rightarrow |\bar{x}_s - \bar{y}_s| = |\bar{x}'_s - \bar{y}'_s|, \text{ and} \end{aligned}$$

$$\forall m \text{ IOb}(m) \rightarrow \exists p \text{ Ph}(p) \wedge W(m, p, 0, 0, 0, 0) \wedge W(m, p, 1, 0, 0, 1).$$

Let us see how AxSm states that “all inertial observers use the same units of measurements.” That “the speed of light is 1” (besides that the speed of light is nonzero) means only that observers are using units measuring time distances compatible with the units measuring spatial distances, such as light years or light seconds. The first part of AxSm means that different observers use the same unit measuring spatial distances. This is so because if two events are simultaneous for both observers, they can measure their spatial distance and the outcome of their measurements are the same iff the two observers are using the same units to measure spatial distances.

Our axiom system for special relativity contains these 5 axioms only:

$$\text{SpecRel} \stackrel{d}{=} \{\text{AxFd}, \text{AxPh}, \text{AxEv}, \text{AxSf}, \text{AxSm}\}.$$

In an axiom system, the axioms are the “price” we pay, and the theorems are the “goods” we get for them. Therefore, we strive for putting only simple, transparent, easy-to-believe statements in our axiom systems. We want to get all the hard-to-believe predictions as theorems. For example, we prove from SpecRel that it is impossible for inertial observers to move faster than light relative to each other (“No FTL travel” for science fiction fans). In the following,  $\vdash$  means logical derivability.

**Theorem 5.1.** (no faster than light inertial observers)

$$\text{SpecRel} \vdash \forall mk\bar{x}\bar{y} \quad \text{IOb}(m) \wedge \text{IOb}(k) \\ \wedge W(m, k, \bar{x}) \wedge W(m, k, \bar{y}) \wedge \bar{x} \neq \bar{y} \rightarrow |\bar{y}_s - \bar{x}_s| < |y_t - x_t|.$$

For a geometrical proof of Thm.5.1, see (Andréka *et al.* 2010).

In relativity theory we are often interested in comparing the worldviews of different observers. So we introduce the worldview transformation between observers  $m$  and  $k$  as the following binary relation:

$$w_{mk}(\bar{x}, \bar{y}) \stackrel{d}{\iff} \text{ev}_m(\bar{x}) = \text{ev}_k(\bar{y}).$$

By Thm.5.2, the worldview transformations between inertial observers in the models of SpecRel are Poincaré transformations, i.e., transformations which preserve the so-called Minkowski-distance  $(y_t - x_t)^2 - |\bar{y}_s - \bar{x}_s|^2$  of  $d$ -tuples  $\bar{y}, \bar{x}$ . For the definition, we refer to (d’Inverno 1992, 110.) or (Misner *et al.* 1973, 66–69.).

**Theorem 5.2.**

$$\text{SpecRel} \vdash \forall m, k \quad \text{IOb}(m) \wedge \text{IOb}(k) \rightarrow w_{mk} \text{ is a Poincaré transformation.}$$

For the proof of Thm.5.2, see (Andréka *et al.* 2007, Thm.11.10, 640.) or (Székely 2009, Thm.3.2.2, 22.). By Thm.5.2, all predictions of special relativity, such as “moving clocks slow down,” are provable from SpecRel. For details, see, e.g., (Andréka *et al.* 2006, §1), (Andréka *et al.* 2007, §2), (Andréka *et al.* 2002, §2.5).

## 6 LOGICAL ANALYSIS

Let us illustrate here by a simple example what we mean by logical analysis of a theory. In AxEv we have assumed that all observers see the same (possibly infinite) meetings of bodies. Let us try to weaken AxEv to an axiom assuming something similar but only for finite meetings of bodies. A natural candidate is one of the following finite approximations of AxEv:

AxMeet<sub>n</sub>: All inertial observers see the same  $n$ -meetings of bodies:

$$\forall mkb_1 \dots b_n \bar{x} \quad \text{IOb}(m) \wedge \text{IOb}(k) \wedge W(m, b_1, \bar{x}) \wedge \dots \wedge W(m, b_n, \bar{x}) \\ \rightarrow \exists \bar{y} \quad W(k, b_1, \bar{y}) \wedge \dots \wedge W(k, b_n, \bar{y}).$$

For example, AxMeet<sub>1</sub> means only that inertial observers see the same bodies. Let us also introduce axiom scheme Meet<sub>ω</sub> as the collection of all the axioms AxMeet<sub>n</sub>. By Prop.6.1, AxMeet<sub>n</sub> is strictly weaker assumption than AxMeet<sub>n+1</sub> and AxEv is strictly stronger than all the axioms of Meet<sub>ω</sub> together.

**Proposition 6.1.**

$$\text{AxEv} \vdash \text{AxMeet}_{n+1} \vdash \text{AxMeet}_n \quad (1)$$

$$\text{AxMeet}_n \not\vdash \text{AxMeet}_{n+1} \quad (2)$$

$$\text{Meet}_\omega \not\vdash \text{AxEv} \quad (3)$$

*Proof.* Item (1) follows easily by the formulations of the axioms.

To prove Item (2), we are going to construct a model of  $\text{AxMeet}_n$  in which  $\text{AxMeet}_{n+1}$  is not valid. Let  $Q = \{0, 1, \dots, n\}$ ,  $B = \{b_i : i \leq n\}$ . Let all the bodies be inertial observers. Let  $b_0$  see all the bodies in  $\langle 0, \dots, 0 \rangle$  and none of them in any other coordinate points, i.e., let  $W(b_0, b_i, \bar{x})$  hold iff  $\bar{x} = \langle 0, \dots, 0 \rangle$ ; and for all  $k \neq 0$  let  $b_k$  see all the bodies but  $b_i$  at coordinate points  $\langle i, \dots, i \rangle$  for all  $i \leq n$ , i.e., let  $W(b_k, b_i, \bar{x})$  hold iff  $\bar{x} = \langle j, \dots, j \rangle$  and  $i \neq j$ . In this model, all inertial observers see all the possible  $n$ -meetings. So  $\text{AxMeet}_n$  is valid in this model. However, the only inertial observer who sees the  $n + 1$ -meeting  $\{b_0, \dots, b_n\}$  is  $b_0$ . So  $\text{AxMeet}_{n+1}$  is not valid in this model.

We are going to prove Item (3) by a similar model construction. The only difference is that now  $Q$  will be infinite. For simplicity, let  $Q$  be the set of natural numbers. Let all the other parts of the model be defined in the same way. Now all the inertial observers see all the possible  $n$ -meetings of the bodies for all natural numbers  $n$ . So  $\text{AxMeet}_n$  is valid in this model for all natural number  $n$ . Hence  $\text{Meet}_\omega$  is valid in this model. However, only  $b_0$  sees the event  $\{b_1, b_2, \dots\}$ . So  $\text{AxEv}$  is not valid in this model.  $\square$

Now we will use that there are no stationary (i.e., motionless) light signals. So let us formalize this statement.

$\text{Ax}(c \neq 0)$ : Inertial observers do not see stationary light signals.

$$\forall mp\bar{x}\bar{y} \quad \text{IOb}(m) \wedge \text{Ph}(p) \wedge W(m, p, \bar{x}) \wedge W(m, p, \bar{y}) \wedge x_t \neq y_t \rightarrow \bar{x}_s \neq \bar{y}_s.$$

**Proposition 6.2.**

$$\text{AxMeet}_3, \text{AxFd}, \text{AxPh}, \text{Ax}(c \neq 0) \vdash \text{AxEv} \quad (4)$$

$$\text{AxMeet}_2, \text{AxFd}, \text{AxPh}, \text{Ax}(c \neq 0) \not\vdash \text{AxEv} \quad (5)$$

$$\text{Meet}_\omega, \text{AxFd}, \text{AxPh} \not\vdash \text{AxEv} \quad (6)$$

*Proof.* First let us make some general observations. By  $\text{AxFd}$ , there is no nondegenerate triangle in  $Q^d$  whose sides are of slope  $c$ . This is clear if  $c = 0$ ; and in the case  $c \neq 0$ , this can be shown by contradiction using the fact that the vertical projection of a triangle of this kind is a triangle whose one side is the sum of the other two sides. Therefore,  $\text{AxFd}$  and  $\text{AxPh}$  together imply that any

inertial observer  $m$  sees the events in which a particular photon participates on a line of slope  $c_m$ .

By AxFd, AxPh and Ax( $c \neq 0$ ), every inertial observer  $m$  sees different meetings of photons at different coordinate points. This is so since (by AxFd) for every pair of points there is a line of slope  $c_m \neq 0$  containing only one of the points. Hence, by AxPh, there is a photon seen by  $m$  only at one of the two coordinate points.

Let us now prove Item (4). Let  $m$  and  $k$  be inertial observers and let  $\bar{x}$  be a coordinate point. To prove AxEv, we have to find a coordinate point  $\bar{x}'$  such that  $ev_m(\bar{x}) = ev_k(\bar{x}')$ . To find this  $\bar{x}'$ , let  $\bar{y} = \langle x_1 + c_m, x_2, \dots, x_{d-1}, x_t + 1 \rangle$ ,  $\bar{z} = \langle x_1 - c_m, x_2, \dots, x_{d-1}, x_t + 1 \rangle$  and  $\bar{w} = \langle x_1, \dots, x_{d-1}, x_t + 2 \rangle$ , see Fig. 1.

By AxPh, there are photons  $p_1$ ,  $p_2$  and  $p_3$  such that  $p_1, p_2 \in ev_m(\bar{x})$ ,  $p_2, p_3 \in ev_m(\bar{y})$ ,  $p_1 \in ev_m(\bar{z})$  and  $p_3 \in ev_m(\bar{w})$ . Since  $m$  sees every photon on a line of slope  $c_m$ , he sees the meeting of  $p_1$  and  $p_2$  only at  $\bar{x}$  and does not see the meeting of  $p_1$  and  $p_3$ .

Since AxMeet<sub>3</sub> implies AxMeet<sub>2</sub>,  $k$  sees the same meetings of pairs of photons. So there is a  $\bar{x}'$  where  $k$  sees  $p_1$  and  $p_2$  meet.  $\bar{x}'$  is the only point where  $k$  sees both  $p_1$  and  $p_2$ . This is so because  $k$  sees different meetings of photons at different points but sees the same 3-meetings as  $m$ . So if there were another point, say  $\bar{x}''$ , where  $k$  sees  $p_1$  and  $p_2$ , there were photons  $p' \in ev_k(\bar{x}')$  and  $p'' \in ev_k(\bar{x}'')$  such that  $p' \notin ev_k(\bar{x}'')$ ,  $p'' \notin ev_k(\bar{x}')$  and  $k$  does not see the meeting of  $p'$  and  $p''$ . By axiom AxMeet<sub>3</sub>  $m$  has to see the meetings  $\{p_1, p_2, p'\}$  and  $\{p_1, p_2, p''\}$ . The only point where  $m$  can see these meetings is  $\bar{x}$  since  $\bar{x}$  the only point where  $m$  sees  $p_1$  and  $p_2$  meet. Therefore  $m$  sees the meeting of  $p'$  and  $p''$  at  $\bar{x}$ . Thus, by AxMeet<sub>3</sub>,  $k$  also has to see the meeting of  $p'$  and  $p''$ , but  $k$  does not see it. Hence  $\bar{x}'$  is the only point where  $k$  sees both  $p_1$  and  $p_2$ .

Let  $b$  be a body such that  $W(m, b, \bar{x})$ . By AxMeet<sub>3</sub>,  $k$  has to see the meeting of  $p_1$ ,  $p_2$  and  $b$ . This point has to be  $\bar{x}'$  since the only point where  $p_1$  and  $p_2$  meet is  $\bar{x}'$ . Since  $b$  was an arbitrary body, we have  $ev_m(\bar{x}) \subseteq ev_k(\bar{x}')$ . The same argument shows that  $ev_k(\bar{x}') \subseteq ev_m(\bar{x})$ . So  $ev_m(\bar{x}) = ev_k(\bar{x}')$  as desired.

We are going to prove Item (5), by constructing a model. Let  $\langle Q, +, \cdot \rangle$  be the field of real numbers. Let us denote the set of natural numbers by  $\omega$ . Let  $B = \{m, k\} \cup \{b_i : i \in \omega\} \cup \{p : p \text{ is a line of slope } 1\}$ . Let  $m$  and  $k$  be all the inertial observers and let the lines of slope 1 be all the photons. Let  $m$  and  $k$  see the photon  $p$  at coordinate point  $\bar{x}$  iff  $\bar{x} \in p$ . Let  $m$  see all the bodies  $b_i$  at  $\bar{x}$  iff  $x_t = 0$ . Let  $k$  see all the bodies  $b_0, \dots, b_n, \dots$  but  $b_i$  at  $\bar{x}$  iff  $x_t = i$  (i.e., iff  $\bar{x}$  is in the horizontal hyperplane  $\{\bar{y} \in Q^d : y_t = i\}$ ).<sup>7</sup> It is straightforward from this construction that axioms AxFd, AxPh and Ax( $c \neq 0$ ) are valid in this model. Since every line of slope 1 intersects every horizontal hyperplane,  $m$  and

<sup>7</sup>If  $d = 2$ , vertical lines can be used instead of horizontal hyperplanes, which gives a counterexample with bodies having more natural properties.

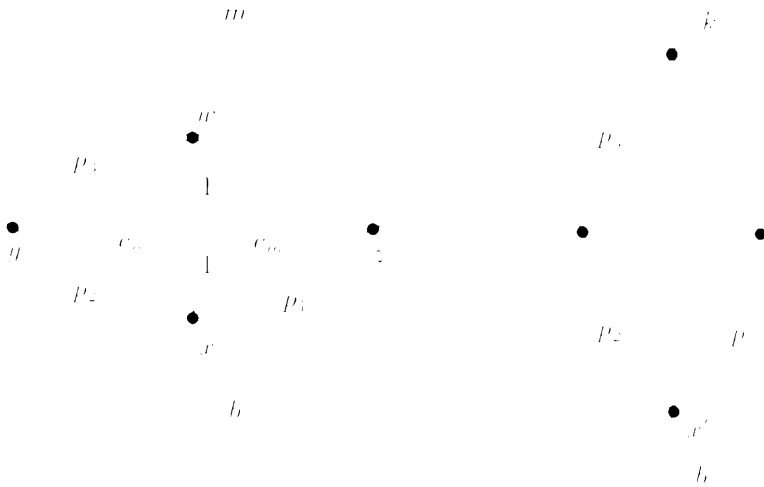


Figure 1.

$k$  see the same 2-meetings of bodies. Hence  $\text{AxMeet}_2$  is also valid in this model. However, the only inertial observer who sees the meeting  $\{b_i : i \in \omega\}$  is  $m$ . So  $\text{AxEv}$  is not valid in this model.

We prove Item (6) by a similar construction. The only difference is that now the set of bodies is  $B = \{m, k\} \cup \{b_i : i \in \omega\} \cup \{p : p \text{ is a vertical line}\}$ ; and the photons are the vertical lines. It is straightforward from the construction that axioms  $\text{AxFd}$ ,  $\text{AxPh}$  are valid in this model ( $c = 0$ ). Since every vertical line intersects every horizontal hyperplane,  $m$  and  $k$  see the same  $n$ -meetings of bodies. Hence  $\text{Meet}_\omega$  is also valid in this model. However, only  $m$  sees the meeting  $\{b_i : i \in \omega\}$ . So  $\text{AxEv}$  is not valid in this model.  $\square$

Prop.6.2 shows that a price to weaken axiom  $\text{AxEv}$  to  $\text{AxMeet}_3$  is to assume that there are no stationary light signals. Since  $\text{AxSm}$  contains this assumption, we can simply replace  $\text{AxEv}$  with  $\text{AxMeet}_3$  in  $\text{SpecRel}$ . A natural continuation of this investigation can be a search for assumptions that allow us to weaken  $\text{AxMeet}_3$  to  $\text{AxMeet}_2$ . A possible candidate is that bodies move along straight lines and the dimension  $d$  is at least 3. The proof of Item (5) shows that assuming only that bodies move along straight lines is not enough, if  $d = 2$ .

We have several similar investigations on the logical connections of axioms and predictions, see, e.g., (Andréka & al. 2008), (Székely 2009, §5) on dynamics, (Madarász & al. 2006), (Székely 2009, §4, §7), (Székely 2010b) on twin paradox, (Andréka & al. 2002) on kinematics, time-dilation and length-contraction, twin paradox, etc.

## 7 ACCELERATED OBSERVERS

In SpecRel we restricted our attention to inertial observers. It is a natural idea to generalize the theory by including accelerated observers as well. It is explained in the classic textbook (Misner & *al.* 1973, 163–165.) that the study of accelerated observers is a natural first step (from special relativity) towards general relativity.

We have not introduced the concept of observers as a basic one because it can be defined as follows: an *observer* is nothing other than a body who “observes” (coordinatizes) some other bodies somewhere, this property can be captured by the following formula of our language:

$$\text{Ob}(m) \stackrel{d}{\iff} \exists b\bar{x} W(m, b, \bar{x}).$$

Our key axiom about accelerated observers is the following:

**AxCmv:** At each moment of his life, every accelerated observer sees (coordinatizes) the nearby world for a short while in the same way as an inertial observer does.

For formulation of AxCmv in our FOI language, see (Madarász & *al.* 2006), (Székely 2009) or (Andréka & *al.* 2010).

Axiom AxCmv ties the behavior of accelerated observers to those of inertial ones. Justification of this axiom is given by experiments. We call two observers *co-moving* at an event if they “see the nearby world for a short while in the same way” at the event. By this notion AxCmv says that at each event of an observer’s life, he has a co-moving inertial observer. We can think of a dropped spacepod as a co-moving inertial observer of an accelerated spaceship (at the event of dropping). Or, if a spaceship switches off its engines, it will move on as a co-moving inertial spaceship would.

Our next two axioms ensure that the worldviews of accelerated observers are big enough. They are generalized versions of the corresponding axioms for inertial observers, but now postulated for all observers.

**AxEv<sup>-</sup>:** If  $m$  sees  $k$  in an event, then  $k$  cannot deny it:

$$\forall m, k \in \text{Ob} \ W(m, k, \bar{x}) \rightarrow \exists \bar{y} \ \text{ev}_m(\bar{x}) = \text{ev}_k(\bar{y}).$$

**AxSf<sup>-</sup>:** Any observer sees himself in an interval of the time axis:

$$\begin{aligned} &\forall m \in \text{Ob} \ \forall \bar{x} \ W(m, m, \bar{x}) \rightarrow x_1 = x_2 = x_3 = 0 \quad \text{and} \\ &\forall \bar{x}\bar{y} \ W(m, m, \bar{y}) \wedge W(m, m, \bar{x}) \rightarrow \forall t \ x_t < t < y_t \rightarrow W(m, m, 0, 0, t). \end{aligned}$$

Our last two axioms will ensure that the worldlines of accelerated observers are “tame” enough, e.g., they have velocities at each moment. In SpecRel,



the worldview transformations between inertial observers are affine maps, the next axiom will state that the worldview transformations between accelerated observers are approximately affine, wherever they are defined.

AxDf: The worldview transformations have linear approximations at each point of their domain (i.e., they are differentiable).

For a precise formalization of AxDf, see, e.g., (Andréka & al. 2010).

We note that AxDf implies that the worldview transformations are functions with open domains. However, if the numberline has gaps, still there can be crazy motions. Our last assumption is an axiom scheme supplementing AxDf by excluding these gaps.

Cont: Every definable, bounded and nonempty subset of  $Q$  has a supremum (i.e., least upper bound).

In Cont “definable” means “definable in the language of AccRel, parametrically.” For a precise formulation of Cont, see (Madarász & al. 2006, 692.) or (Székely 2009, §10.1). Cont is a “mathematical axiom” in spirit. It is Tarski’s FOI version of Hilbert’s continuity axiom in his axiomatization of geometry, see (Goldblatt 2004, 61–162.), fitted to the language of AccRel. When  $Q$  is the field of real numbers, Cont is automatically true.

That Cont requires the existence of supremum only for sets definable in the language of AccRel instead of every set, is important not only because by this trick we can keep our theory within FOI (which is crucial in a foundational work), but also because it makes this postulate closer to the the physical/empirical level. The latter is true because Cont does not speak about “any fancy subset” of the quantities, just those “physically meaningful” sets which can be defined in the language of our (physical) theory.

Adding this 5 axioms to SpecRel, we get an axiom system for accelerated observers:

$$\text{AccRel} \stackrel{d}{=} \text{SpecRel} \cup \{ \text{AxCmv}, \text{AxEv}^-, \text{AxSf}^-, \text{AxDf} \} \cup \text{Cont}.$$

As an example we show that the so-called *twin paradox* can be naturally formulated and analyzed logically in AccRel. Our axiomatic approach also makes it possible to analyze the details of the twin paradox (e.g., who sees what, when) with the clarity of logic, see (Andréka & al. 2002, 139–150.) for part of such an analysis.

According to the twin paradox, if a twin makes a journey into space (accelerates), he will return to find that he has aged less than his twin brother who stayed at home (did not accelerate). We formulate the twin paradox in our FOI language as follows.

TWP: Every inertial observer  $m$  measures at least as much time as any other observer  $k$  between any two events  $e_1$  and  $e_2$  in which they meet;

and they measure the same time iff they have encountered the very same events between  $e_1$  and  $e_2$ :

$$\begin{aligned} \forall m \in \text{IOb} \ \forall k \in \text{Ob} \ \forall \bar{x} \bar{x}' \bar{y} \bar{y}' \ x_t < y_t \wedge x'_t < y'_t \wedge \\ m, k \in \text{ev}_m(\bar{x}) = \text{ev}_k(\bar{x}') \wedge m, k \in \text{ev}_m(\bar{y}) = \text{ev}_k(\bar{y}') \rightarrow y'_t - x'_t \leq y_t - x_t \\ \wedge (y'_t - x'_t = y_t - x_t \leftrightarrow \text{enc}_m(\bar{x}, \bar{y}) = \text{enc}_k(\bar{y}', \bar{y}')), \end{aligned}$$

where  $\text{enc}_m(\bar{x}, \bar{y}) = \{\text{ev}_m(\bar{z}) : W(m, m, \bar{z}) \wedge x_t \leq z_t \leq y_t\}$ .

### Theorem 7.1.

$$\text{AccRel} \vdash \text{TwP} \quad (7)$$

$$\text{AccRel} - \text{AxDf} \vdash \text{TwP} \quad (8)$$

$$\text{AccRel} - \text{Cont} \not\vdash \text{TwP} \quad (9)$$

$$\text{Th}(\mathbb{R}) \cup \text{AccRel} - \text{Cont} \not\vdash \text{TwP} \quad (10)$$

For the proof of Thm.7.1, see (Madarász & al. 2006) or (Székely 2009, §7).

Item (10) of Thm.7.1 states that Cont cannot be replaced with the whole FOL theory of real numbers in AccRel if we do not want to loose TwP from its consequences.

Our theory AccRel is also strong enough to predict the gravitational time-dilation effect of general relativity via Einstein's equivalence principle, see (Madarász & al. 2007), (Székely 2009).

## 8 GENERAL RELATIVITY

Our theory of accelerated observers AccRel speaks about two kinds of observers, inertial and accelerated ones. Some axioms are postulated for inertial observers only, some apply to all observers. We get an axiom system GenRel for general relativity by stating the axioms of AccRel in a generalized form in which they are postulated for all observers, inertial and accelerated ones equally. In other words, we will change all axioms of AccRel in the same spirit as AxSf<sup>-</sup> and AxEv<sup>-</sup> were obtained from AxSf and AxEv, respectively. This kind of change AccRel  $\mapsto$  GenRel can be regarded as a “democratic revolution” with the slogan “all observers should be equivalent, the same laws should apply to all of them.” Here “law” translates as “axiom.” This idea originates with Einstein (see his book (Einstein 1921/2006, Part II, ch.18)).

For simplicity, we will use an equivalent version of the symmetry axiom AxSm (see (Andréka & al. 2002, Thm.2.8.17(ii), 138.) or (Székely 2009, Thm.3.1.4, 21.)), and we will require the speed of photons to be 1 in AxPh<sup>-</sup> (as opposed to requiring it in AxSm<sup>-</sup>).

AxPh<sup>-</sup>: The velocity of photons an observer “meets” is 1 when they meet, and it is possible to send out a photon in each direction where the observer stands.

AxSm<sup>-</sup>: Meeting observers see each other’s clocks slow down with the same rate.

For a precise formulation of these axioms, see (Andréka & *al.* 2010), (Székely 2009).

We introduce an axiom system for general relativity as the collection of the following axioms:

$$\text{GenRel} \stackrel{d}{=} \{ \text{AxFd}, \text{AxPh}^-, \text{AxEv}^-, \text{AxSf}^-, \text{AxSm}^-, \text{AxDf} \} \cup \text{Cont.}$$

Axiom system GenRel contains basically the same axioms as SpecRel, the difference is that they are assumed only locally but for all the observers.

Thm.8.1 below states that the models of GenRel are exactly the spacetimes of usual general relativity. For the notion of a Lorentzian manifold we refer to (d’Inverno 1992, 55.), (Misner & *al.* 1973, 241.) and (Andréka & *al.* 2007, sec.3.2).

**Theorem 8.1** (Completeness theorem). *GenRel is complete with respect to its standard models, i.e., with respect to Lorentzian Manifolds over real closed fields.*

This theorem can be regarded as a completeness theorem in the following sense. Let us consider Lorentzian manifolds as intended models of GenRel. How can we do that? We give a method for constructing a model of GenRel from each Lorentzian manifold; and conversely, we show that each model of GenRel is obtained this way from a Lorentzian manifold. After this is elaborated, we have defined what we mean by a formula  $\varphi$  in the language of GenRel being valid in a Lorentzian manifold. Then completeness means that for any formula  $\varphi$  in the language of GenRel, we have  $\text{GenRel} \vdash \varphi$  iff  $\varphi$  is valid in all Lorentzian manifolds over real closed fields. This is completely analogous to the way in which Minkowskian spacetimes were regarded as intended models of SpecRel in the completeness theorem of SpecRel, see (Andréka & *al.* 2007, Thm.11.28, 681.) and (Madarász 2002, §4).

We call the worldline of an observer *timelike geodesic*, if each of its points has a neighborhood within which this observer “maximizes measured time (wrist-watch time)” between any two encountered events. For formalization of this concept in our FOI language, see, e.g., (Andréka & *al.* 2010).

According to the definition above, if there are only a few observers, then it is not a big deal that a worldline is a time-like geodesic (it is easy to be maximal if there are only a few to be compared to). To generate a real competition for the rank of having a timelike geodesic worldline, we postulate the existence of many observers by the following axiom scheme of comprehension.

Compr: For any parametrically definable timelike curve in any observers worldview, there is another observer whose worldline is the range of this curve.

A precise formulation of Compr can be obtained from that of its variant in (Andreka & al. 2007, 679.).

An axiom schema Compr guarantees that our definition of a geodesic coincides with that in the literature on Lorentzian manifolds. Therefore we also introduce the following theory:

$$\text{GenRel}^+ \stackrel{d}{=} \text{GenRel} \cup \text{Compr}.$$

So in our theory  $\text{GenRel}^+$ , our concept of timelike geodesic coincides with the standard concept in the literature on general relativity. All the other key concepts of general relativity, such as curvature or Riemannian tensor field, are definable from timelike geodesics. Therefore we can treat all these concepts (including the concept of metric tensor field) in our theory  $\text{GenRel}^+$  in a natural way.

In general relativity, Einstein's field equations (EFE) provide the connection between the geometry of spacetime and the energy-matter distribution (given by the energy-momentum tensor field). Since in  $\text{GenRel}^+$  all the geometric concepts of spacetime are definable, we can use Einstein's equation as a definition of the energy-momentum tensor, see, e.g., (Benda 2008) or (d'Inverno 1992, §13.1, 169.), or we can extend the language of  $\text{GenRel}^+$  with the concept of energy-momentum tensor and assume Einstein's equations as axioms. As long as we do not assume anything more of the energy-momentum tensor than its connection to the geometry described by Einstein's equations, there is no real difference in these two approaches. In both approaches, we can add extra conditions about the energy-momentum tensor to our theory, e.g., the dominant energy condition or, e.g., that the spacetimes are vacuum solutions.

## 9 CAN PHYSICS GIVE FEEDBACK TO LOGIC?

There is observational evidence suggesting that in our physical universe there exist regions supporting potential non-Turing computations. Namely, it is possible to design a physical device in relativistic spacetime which can compute a non-Turing computable task, e.g., which can decide whether ZF set theory is consistent. This empirical evidence is making the theory of hypercomputation more interesting and gives new challenges to the physical Church Thesis, see, e.g., (Andréka & al. 2009).

These new challenges do more than simply providing a further connection between logic and spacetime theories; they also motivate the need for logical understanding of spacetime theories.

## 10 CONCLUDING REMARKS

We have axiomatized both special and general relativity in FOL. Moreover, via our theory AccRel, we have axiomatized general relativity so that each of its axioms can be traced back to its roots in the axioms of special relativity. Axiomatization is not our final goal. It is merely an important first step toward logical and conceptual analysis. We are only at the beginning of our ambitious project.\*

## REFERENCES

- Andréka, H., J. X. Madarász, and I. Németi, with contributions from A. Andai, G. Sági, I. Sain and Cs. Tóke, 2002, *On the Logical Structure of Relativity Theories*. Research report. Budapest, Alfréd Rényi Institute of Mathematics. <http://www.renyi.hu/pub/algebraic-logic/Contents.html>.
- Andréka, H., J. X. Madarász, and I. Németi, 2006, Logical axiomatizations of space-time. Samples from the literature. In A. Prékopa, & al. (eds.) *Non-Euclidean Geometries*. Berlin, Springer, 155–185.
- Andréka, H., J. X. Madarász, and I. Németi, 2007, Logic of space-time and relativity theory. In M. Aiello, & al. (eds.), *Handbook of Spatial Logics*. Berlin, Springer, 607–711.
- Andréka, H., J. X. Madarász, I. Németi, and G. Székely, 2008, Axiomatizing relativistic dynamics without conservation postulates. *Studia Logica* 89, 163–186.
- Andréka, H., I. Németi, and P. Németi, 2009, General relativistic hypercomputing and foundation of mathematics. *Nat. Comp.* 8, 499–516.
- Andréka, H., J. X. Madarász, I. Németi, and G. Székely, 2010, A logic road from special relativity to general relativity. *Synthese*, submitted.
- Ax, J., 1978, The elementary foundations of spacetime. *Found. Phys.* 8, 507–546.
- Benda, T., 2008, A formal construction of the spacetime manifold. *J. Philos Logic* 37, 441–478.
- Chang, C. C., and H. J. Keisler, 1990, *Model theory*. Amsterdam, North-Holland.
- d’Inverno, R., 1992, *Introducing Einstein’s relativity*. Oxford, Oxford Univ. Press.
- Einstein, A., 1905/1952, Zur Elektrodynamik bewegter Körper. *Annalen der Physik* 17, 891–921. English translation in A. Einstein, *The principle of Relativity*. Mineola (NY), Dover.
- Einstein, A., 1921/2006, *Relativity. The Special and the General Theory*. London, Penguin Classics. Translated by W. Lawson.
- Friedman, H., 2004a, On foundational thinking 1. Posting in FOM (Foundations of Mathematics) Archives [www.cs.nyu.edu](http://www.cs.nyu.edu) (Jan. 20, 2004).
- Friedman, H., 2004b, On foundations of special relativistic kinematics 1. Posting in FOM (Foundations of Mathematics) Archives [www.cs.nyu.edu](http://www.cs.nyu.edu) (Jan. 21, 2004).
- Goldblatt, R., 2004, *Orthogonality and spacetime geometry*. Berlin, Springer.
- Gömöri, M., and L. E. Szabó, 2010, Is the relativity principle consistent with electrodynamics? Towards a logico-empiricist reconstruction of a physical theory. arXiv:0912.4388v3.
- Guts, A. K., 1982, The axiomatic theory of relativity. *Russ. Math. Surv.* 37, 41–89.
- Hershbach, D., 2008, Einstein as a student. In P. L. Galison & al. (eds.) *Einstein for the 21st century*. Princeton, Princeton Univ. Press, 217–238.
- Hirsch, R., 2009, Relativity and modal logic. *Hungarian Philosophical Review*, this issue.

\*This research is supported by the Hungarian Scientific Research Fund for basic research grant No. T81188, as well as by a Bolyai grant for J. X. Madarász.

- Madarász, J. X., 2002, *Logic and Relativity: in the Light of Definability Theory*. PhD thesis, Budapest, Eötvös Loránd Univ.
- Madarász, J. X., I. Németi, and G. Székely, 2006, Twin paradox and the logical foundation of relativity theory. *Found. Phys.* 36, 681–714.
- Madarász, J. X., I. Németi, and G. Székely, 2007, First-order logic foundation of relativity theories. In D. Gabbay, & al. (eds.), *Mathematical Problems from Applied Logic II*. Berlin, Springer, 217–252.
- Misner, C. W., K. S. Thorne, and J. A. Wheeler, 1973, *Gravitation*. New York, W. H. Freeman and Co.
- Rosinger, E. E., 2008, *Two Essays on the Archimedean versus Non-Archimedean Debate*. arXiv:0809.4509v3.
- Rosinger, E. E., 2009, *Special Relativity in Reduced Power Algebras*. arXiv:0903.0296v1.
- Schutz, J. W., 1973, *Foundations of Special Relativity: Kinematic Axioms for Minkowski Space-Time*. Berlin, Springer.
- Suppes, P., 1968, The desirability of formalization in science. *J. Philos.* 27, 651–664.
- Szabó, L. E., 2009, Empirical Foundation of Space and Time. In M. Suárez, & al. (eds.), *EPSA07: Launch of the European Philosophy of Science Association*. Berlin, Springer.
- Székely, G., 2009, *First-Order Logic Investigation of Relativity Theory with an Emphasis on Accelerated Observers*. PhD thesis, Budapest, Eötvös Loránd Univ.
- Székely, G., 2010a, On why-questions in physics. In F. Stadler & al., (ed.), *Wiener Kreis und Ungarn*. Berlin, Springer, to appear.
- Székely, G., 2010b, A geometrical characterization of the twin paradox and its variants. *Studia Logica*, online-first.
- Tarski, A., 1959, What is elementary geometry? In L. Henkin, & al. (eds.) *The Axiomatic Method. With Special Reference to Geometry and Physics*. Amsterdam, North-Holland, 16–29.
- Väänänen, J., 2001, Second-order logic and foundations of mathematics. *Bull. Symb. Log.* 7, 504–520.
- Wolenski, J., 2004, First-order logic: (philosophical) pro and contra. In V. F. Hendricks & al. (eds.), *First-Order Logic Revisited*. Berlin, Logos, 369–398.

## Modal Logic and Relativity

**Abstract.** We argue that modal logic is the natural logic to use to reason about relativity theory. We define a complete modal axiomatisation of the kinematics of special relativity theory.

Relativity Theory, in its most general sense, rejects the notion of absolute space. In Relativity Theory statements such as “this rod is one meter long” are frowned upon, rather we should say “this rod is one meter long when measured in this frame of reference”. When devising a logic to reason about relativity theory, it is therefore natural to adopt a modal logic where every statement has an implicit ‘point of view’. In that sense, a modal logic might be more true to the subject matter of relativity theory. We do not claim that such an approach will in itself provide new technical results in relativity theory, but apparent paradoxes and other conceptual difficulties with relativity theory might be more easily avoided in a modal setting. A second motivation for modal logic is that the complexity of reasoning in a modal logic can often be lower than with first-order logic. Thirdly, the current article is only an initial step and we believe that modal logic should be used to reason about general relativity where it is even more important to work in a local framework. In this article we will consider how a modal logic for special relativity theory might be devised.

A fundamental concept in relativity theory is that of the observation. For Einstein, and his group of followers in the *Vienna Circle*, the precise nature of an observation was of great importance. In Einstein’s original paper on Special Relativity (Einstein 1905), he takes care to clarify the meaning of words like “time”, “simultaneous”, “length”, etc. by replacing them by statements concerning observers and observations. Later, *Logical Positivists* formulated the *verification principle*, which stated that a proposition could be held to be true to the extent that it could be tested by experiments and observations. The critical thing about observations in relativity theory is that what is observed depends

not only on the event but on the observer too. Later, we will define a modal logic with a Kripke semantics in which each observer is a Kripke world.

One intriguing feature of observers is that they act both as subjects and as objects—they see, but they can also be seen. In Einstein’s original paper, he refers to an observer as “the man at the railway-carriage window”, suggesting a point-like body moving through space. In most presentations of relativity theory, an observer is a point-like body with its own world line in 4D spacetime. This tells us that when we see an observer, he looks one-dimensional. We see the observer’s time axis, but we do not see his space axes. However, from his point of view, an observer can see various events taking place at various spacetime points distributed throughout the four dimensions of spacetime. Indeed, when we think of an observer as a subject, it is better not to think of an individual point-like body, but a whole team of colleagues arranged in a grid in three-dimensional space, not moving relative to each other over time, who send messages to each other (or perhaps to a central control centre) about their immediate observations. But as we mentioned, only one member of the team of observers is directly visible to observers from other reference frames. These two aspects of observers are related in (Andréka & *al.* 2010) by an axiom that requires that all observers see themselves at the space origin of their reference frame, at all times, i.e. their world line is the time axis.

Here we consider two different applications of modal logic to special relativity theory: the first approach has to do with modal frame definability, and the second uses model definability. To take a simple and perhaps more familiar example of frame definability, the logic **S4** with axioms  $\{\Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q), \Box p \rightarrow p, \Box p \rightarrow \Box \Box p\}$  and inference rules modus ponens and necessitation defines the set of all validities over Kripke frames whose accessibility relation is reflexive and transitive, that is, **S4** defines the validities of the class of reflexive transitive frames. Furthermore, if these axioms are valid over any Kripke frame  $(W, R)$ , then  $R$  is transitive and reflexive over  $W$ , i.e. **S4** defines the class of reflexive transitive frames. For model definability, if we take the class of reflexive transitive frames as given, the formula  $\Box(p \rightarrow \Diamond \neg p)$ , which is not valid over any frame, defines those models where  $p$  is never ‘eventually true’.

For special relativity, the frame definability approach has already been considered. A Kripke frame may be defined whose Kripke worlds are the points of a four-dimensional Minkowski spacetime and where the accessibility relation is “can send a signal to” and the problem is to define a modal logic that derives all modal formulas valid over frames with this accessibility relation. If we require that signals travel at less than the speed of light and that a signal may be sent from a spacetime point to itself (reflexivity of the accessibility relation) then the set of modal validities is **S4.2** (Goldblatt 1980; Shehtman 1983), the logic obtained by adding the axiom  $\Diamond \Box p \rightarrow \Box \Diamond p$  to the axioms of **S4**. However, **S4.2** does not define this class of four-dimensional models, as the same set of



validities hold over Minkowski spacetime of dimension 2, 3, ..., over Galilean models or even over models with discrete time. S4.2 actually defines the class of all reflexive, transitive and *directed* frames. If you consider an irreflexive accessibility relation “can send a signal to a different point” then modal logic can be more discriminating, however, the purpose of all this is to axiomatise the validities of the given accessibility relation, it does not express the lengths of rods or how they may be transformed when viewed by another observer.

The main focus of the current article is about model definability. We seek a modal logic which is able to express the kinds of observations we might want to make in four-dimensional spacetime, and it should also be able to express the kinds of observations we might expect other observers to make. It can be considered as part of a line of research which seeks to provide simple logical axioms from which the main theorems of relativity theory can be derived. Many such logics have been devised, we make no attempt to survey them here, see (Andréka & *al.* 2006) for an extensive survey. In general, previous attempts to provide a logic for relativity adopt a first-order, or in many cases second-order (e.g. the axiom of continuity in (Schutz 1997)) logic. Consequently they adopt Tarskian semantics. Global variables may be used to denote bodies, field values etc. Instead of the Newtonian statement “Body  $b$  is at  $(x, y, z)$  at time  $t$ ”, the dependence of an observation on the observer is expressed by an observation predicate  $W$ , so that  $W(o, b, x, y, z, t)$  expresses “observer  $o$  sees body  $b$  at  $(x, y, z, t)$ ”.

In a modal logic in which the role of Kripke world is taken by an inertial frame of reference, the dependence on the observer will be suppressed. The modal operator  $\Diamond$  will permit us to transfer from one inertial frame to another. We will take as given a frame in which the accessibility relation is the universal relation. But what language should be adopted to describe the observations made within a single reference frame? In order to keep the presentation fairly general, here we use a two sorted first-order logic with one sort for bodies ( $B$ ) and the other for quantities ( $Q$ ). Variables of sort  $Q$  will be used to record coordinate values. We may use subscripts  $b, q$  for constants, functions and predicates to indicate their sort. The predicate  $See_{q^4b}$  requires four coordinates and one body and tells us that this body is observed at the four coordinate values. We avoid a number of difficulties that sometimes arise with modal first-order languages by requiring that the first-order variables have the same domain at each world of a structure.

## The Language

Variables:	$b, c, \dots$ (type $B$ ), $x, y, z, \dots$ (type $Q$ )
Constants/functions:	$0, 1, +, \times$ (type $Q$ )
Predicates:	$Ph_b, Obs_b, \leq_{qq}, See_{q^4b}$
Formulas:	$\phi ::= Atom \mid \neg\phi \mid (\phi_1 \vee \phi_2) \mid \exists var\phi \mid \Diamond\phi$

The following abbreviations will be useful later.

$$\begin{aligned} \text{vel}(b) = (v_0, v_1, v_2) \quad \text{means} \quad & \exists \bar{x} \forall \bar{y} \\ & [\text{See}(\bar{y}b) \leftrightarrow \exists \lambda \bar{y} = \bar{x} + \lambda \times (v_0, v_1, v_2, 1)] \\ |(x_0, x_1, x_2)| \quad \text{means} \quad & \sqrt{x_0^2 + x_1^2 + x_2^2} \end{aligned}$$

**Structure** As promised, our semantics will be based on Kripke-like structures, where the Kripke worlds are inertial frames of reference. Thus a structure will have the form

$$(W, \beta, F, I, W \times W)$$

where  $W$  is the set of Kripke worlds,  $\beta$  is the set of bodies,  $F$  is the set of quantities,  $I$  interprets variables as elements of  $\beta$  or  $F$  (depending on the sort of the variable),  $0, 1, +, \times, \leq$  as functions/predicates on  $F$ , and  $I$  satisfies  $I(Ph) \cup I(Obs) \subseteq \beta$ ,  $I(\text{See}) \subseteq W \times F^4 \times \beta$ . As we mentioned before, the sets  $\beta, F$  are the same for all worlds (constant domains). Further, all constants, functions and predicates in our language are rigidly designated, except for  $\text{See}_{q^4b}$  which is expected to vary from one world to another. Given such a structure, we may evaluate formulas in the obvious way. Let  $\mathcal{S} = (W, \beta, F, I, W \times W)$  and  $w \in W$ .

$$\begin{aligned} \mathcal{S} \models \text{Obs}(b) & \iff I(b) \in I(\text{Obs}) \\ \mathcal{S} \models t \leq s & \iff (I(t), I(s)) \in I(\leq) \\ \mathcal{S}, w \models \text{See}(x, y, z, t, b) & \iff (w, I(x, y, z, t, b)) \in I(\text{See}) \\ \mathcal{S}, w \models \exists x \phi & \iff (W, \beta, F, I', W \times W), w \models \phi \\ & \quad \text{(some } I' \text{ that agrees with } I \text{ except per-} \\ & \quad \text{haps on } x) \\ \mathcal{S}, w \models \Diamond \phi & \iff \mathcal{S}, v \models \phi \text{ (some } v \in W) \end{aligned}$$

**Axioms for defining class of structures**  $\Box$ - $\forall$ -closure of:

- (1)  $(F, 0, 1, +, \times, \leq)$  is a Real Closed Field, i.e. an ordered field where every non-negative element has a square root (Euclidean) and every polynomial of odd degree has a root.

In many presentations of special relativity, the ordered field is only required to be Euclidean and this suffices for most of our results. Here we assume that the field is a Real Closed Field in order to obtain a decidability result. Note that all Real Closed Fields are elementarily equivalent to the real numbers (Tarski 1951).

- (2) Observers are inertial:  $\text{Obs}(b) \rightarrow \exists v_0 v_1 v_2 (\text{vel}(b) = (v_0, v_1, v_2))$

(3) Speed of light is constant:  $Ph(b) \rightarrow |\text{vel}(b)| = 1$

(4) There is an observer on every 'slow line', there is a photon on every 'fast line':

$$|(v_0, v_1, v_2)| < 1 \rightarrow \exists b(\text{Obs}(b) \wedge \text{See}(\bar{x}, t, b) \wedge \text{vel}(b) = (v_0, v_1, v_2))$$

$$|(v_0, v_1, v_2)| = 1 \rightarrow \exists b(Ph(b) \wedge \text{See}(\bar{x}, t, b) \wedge \text{vel}(b) = (v_0, v_1, v_2))$$

In view of the last axiom, for each observer and for all  $x, y, z, t \in F$ , there are bodies  $b_1, b_2, b_3$  moving on distinct lines that meet uniquely (pairwise and jointly) at  $(x, y, z, t)$ , according to that observer. The triple  $e = (b_1, b_2, b_3)$  is called an event, we may write  $\text{See}(x, y, z, t, e)$  instead of  $\bigwedge_{i=1,2,3} \text{See}(x, y, z, t, b_i)$ .

(5) All observers see the same events:

$$\text{See}(x, y, z, t, e) \rightarrow \Box \exists x, y, z, t \text{ See}(x, y, z, t, e).$$

(6) Symmetry. Let  $e_0, e_1, e'_0, e'_1$  be events.

$$\bigwedge_{i=0,1} (\text{See}(0, 0, 0, i, e_i) \wedge \text{See}(x'_i, y'_i, z'_i, t'_i, e'_i)) \rightarrow$$

$$\Box (\bigwedge_{i=0,1} (\text{See}(0, 0, 0, i, e'_i) \wedge \text{See}(x_i, y_i, z_i, t_i, e_i)) \rightarrow (t'_1 - t'_0 = t_1 - t_0))$$

(7) Isotropy:  $(\bigwedge_{i,j < 4} (|\bar{x}_i - \bar{x}_j| = |\bar{x}'_i - \bar{x}'_j|) \wedge \bigwedge_{i < 4} \text{See}(\bar{x}_i, t, b_i))$

$$\rightarrow \bigwedge_{i < 4} \text{See}(\bar{x}'_i, t', b_i)$$

where  $\bar{x}_i$  is a triple of three spatial coordinates, for  $i < 4$ .

The symmetry axiom implies that any two observers see each other's clocks slow at the same rate. This usefully rules out a situation where one observer measures in seconds while another observer measures in years, it also rules out the situation where one observer measures time going forward while the other measures time going backwards. The isotropy axiom relates to the difference between the one dimensional appearance of observers and the four dimensions that an observer sees. Recall that when we see an observer, we see only his time axis, we do not see the orientation of his spatial axes. The isotropy axiom states that if I can see four bodies at  $\bar{x}_0, \dots, \bar{x}_3$  at time  $t$ , and if the spatial Euclidean distances between the  $\bar{x}_i$  are identical to the distances between the  $\bar{x}'_i$ , then another observer can see the same four bodies at  $\bar{x}'_0, \dots, \bar{x}'_3$  at time  $t'$ . I can transform myself to the second observer by performing an isometry of the spatial coordinates followed by a time translation. Let  $Ax$  be the set of six axioms, just defined.

Having defined the semantics of our language and the axioms for our logic, we now briefly evaluate our system by three criteria: how expressive is the language? are the axioms complete over an appropriate class of structures? what is the complexity of the satisfiability problem for formulas in our language,

over Minkowski structures? As far as expressive power is concerned, it seems that this language is capable of expressing most of the technical statements you find in a textbook on special relativity. For example, given a velocity vector  $\bar{v} = (v_0, v_1, v_2)$  and any formula  $\psi$  we may write  $\Diamond_{\bar{v}}\psi$  for  $\exists b(\text{vel}(b) = \bar{v} \wedge \Diamond(\text{vel}(b) = 0 \wedge \psi))$ , which means “there is a frame moving with velocity  $\bar{v}$  and  $\psi$  is true in that frame”. Our language should be able to express the purely kinematic properties of special relativity. However, our language can only express kinematic statements, we are not able to express properties relating to mass, energy or electric charge, for example.

Next, we assess the deductive strength of our axioms.

**Lemma 1.** *Let  $\mathcal{S} = (W, \beta, F, I, W \times W) \models$  Axioms 1-6 and let  $w, v \in W$ . There is a Poincaré map  $p : F^4 \rightarrow F^4$  (an isometry with respect to Minkowski distance in  $F^4$ ) such that*

$$(1) \quad \mathcal{S}, w \models \text{See}(x, y, z, t, b) \iff \mathcal{S}, v \models \text{See}(p(x, y, z, t))$$

PROOF:

A map  $p : F^4 \rightarrow F^4$  satisfying (1) is uniquely defined, since  $\mathcal{S} \models 4, 5$ . We have to show that  $p$  is a Poincaré map. By axioms 2 and 3, for any  $b \in I(\text{Obs}) \cup I(\text{Ph})$ , the set  $\{(x, y, z, t) : \mathcal{S}, w \models \text{See}(x, y, z, t, b)\}$  is a line of  $F^4$  and by axiom 4 each light line of  $F^4$  is the trace of a photon and each slow line is the trace of an observer. Thus  $p$  maps lines to lines and maps light lines to light lines. By the Alexandrov-Zeeman theorem,  $p$  is a Poincaré transformation followed by a dilation and a field automorphism induced transformation. By axiom 6, the dilation and field induced transformations must be the identity transformations.  $\square$

The isotropy axiom, axiom 7, is needed to complete the proof of the next lemma.

**Lemma 2.** *Let  $\mathcal{S} \models Ax$ ,  $\mathcal{S}' \models Ax$  be two models of our axioms, where  $\mathcal{S} = (W, \beta, F, I, W \times W)$  and  $\mathcal{S}' = (W', \beta', I', F', W' \times W')$ . There are maps  $i : W \rightarrow W'$ ,  $j : \beta \rightarrow \beta'$  and  $k : F \rightarrow F'$  such that  $k$  is an ordered field embedding, and for all  $w \in W$ ,  $b \in \beta$ ,*

$$\mathcal{S}, w \models \text{See}(x, y, z, t, b) \iff \mathcal{S}', i(w) \models \text{See}(k((x), k(y), k(z), k(t), j(b)))$$

Next, we define a standard model  $\mathcal{M}$ . The set of worlds of  $\mathcal{M}$  is the set of Poincaré transformations of  $\mathbb{R}^4$ . If  $p$  is a Poincaré transformation and  $l$  is a line then we write  $p(l)$  for the image of  $l$  under  $p$ , note that  $p(l)$  is always itself a line. The set of bodies of  $\mathcal{M}$  is  $L$ , the set of lines of  $\mathbb{R}^4$  of gradient at most one, observers are the slow lines of gradient strictly less than one and photons are the lines of gradient one. The interpretation  $I_M$  in the standard model is given by

$$(p, x, y, z, t, b) \in I_M(\text{See}) \iff (x, y, z, t) \in p(b)$$

**Theorem 1.**

- (1)  $\mathcal{M}$  is a model of  $Ax$ .
- (2) For any formula  $\phi$  of our language we have

$$\mathcal{M} \models \phi \iff Ax \vdash \phi$$

The first part can easily be proved, simply by verifying that each of our axioms holds in  $\mathcal{M}$ . The second part follows from lemma 2. Thus our axioms are complete for the validities of the standard model.

We now consider the decidability of the following decision problem: Is modal formula  $\phi$  satisfiable over models of  $Ax$ ? If we allow arbitrary bodies to exist in our models, then this problem is undecidable. We may reduce the tiling problem to this satisfiability problem by considering each tile as a body that may be observed at positions with integer coordinates and where the adjacencies are expressed by universally quantified modal formulas. This undecidability arises purely from the first-order part of our language and can be proved without using the modalities. However, if we restrict to bodies whose paths are described by polynomial equations, then the problem becomes decidable. Since we have a universal modality, an arbitrary modal formula  $\phi$  may be converted to a disjunction of clauses of the form  $\Box\psi \wedge \bigwedge_i \Diamond\psi_i$ , where  $\psi, \psi_i$  are non-modal. Such a clause is satisfiable iff  $\psi \wedge \psi_i$  is satisfiable, for each  $i$ . By our assumption about the trajectories of bodies, such a non-modal formula may be translated into a first-order formula in a language with a binary predicate  $\leq$  and functions  $+$ ,  $\times$ . Tarski showed (Tarski 1951; Canny 1988), by elimination of quantifiers, that the satisfiability problem for this language over real closed fields is decidable, although the complexity is rather high (at least double exponential). By imposing restrictions on the use of the first-order connectives in our language, satisfiability problems with lower complexities may be obtained.

**Questions**

- As an alternative to the current exposition, consider a modal logic for special relativity theory where an observer sees a body not merely as a line in four-dimensional space but as a line with a spatial orientation, so at an instance we see a point and three spatial unit vectors.
- Can we define models of special relativity without variables? One approach would be to use a propositional modal logic with S5 modalities for moving in the directions of each of the spatial unit vectors and a temporal modality for moving in time, along with the already included modality for changing reference frame. It is known that this class of frames cannot be finitely axiomatised and the equational theory is undecidable (Hirsch & al. 2002), but here we are more interested in model definability.

- Can we use a similar framework to define a modal logic of general relativity? A Kripke world could be considered as an open neighbourhood in a coordinate system and the whole Kripke frame would correspond to a manifold. An extra modality to transfer to an adjacent neighbourhood would be needed. See (Shapirovski and Shehtman 2005) for useful results on modal logics of regions in Minkowski spacetime.

## REFERENCES

- Andréka, H., Madarász, J. and Németi, I., 2010, *On the Logical Structure of Relativity Theories (Making Relativity Modular, Changeable, and Easy)*. Dordrecht, Kluwer. Provisionally accepted to appear with Kluwer. Draft available at <http://www.math-inst.hu/pub/algebraic-logic/contents.html>.
- Andréka, H., Madarász, J. X. and Németi, I., 2006, Logical axiomatizations of space-time. Samples from the literature. In A. Prékopa, et al. (eds.) *Non-Euclidean Geometries*. Berlin, Springer, 155–185.
- Canny, J., 1988, Some algebraic and geometric computations in pspace. In *Proc. of the 20th SCM symposium on theory of computing*, 460–467.
- Einstein, A., 1905, Zur Elektrodynamik bewegter Körper. *Annalen der Physik* 17, 891–921. English translation available at <http://www.fourmilab.ch/etexts/einstein/specrel/www/>.
- Goldblatt, R., 1980, Diodorean modality in Minkowski space-time. *Studia Logica* 39, 219–236.
- Hirsch, R., Hodkinson, I., and Kurucz, A., 2002, On modal logics between  $K \times K \times K$  and  $S5 \times S5 \times S5$ . *J. Symbolic Logic* 67, 221–234.
- Schutz, J., 1997, *Independent Axioms for Minkowski Space-Time*. Pitman Research Notes in Mathematics. New York, Longman.
- Shehtman, V., 1983, Modal logics of domains on the real plane. *Studia Logica* 64, 63–88.
- Shapirovski, I. and Shehtman, V., 2005, Modal logics of regions and Minkowski spacetime. *Journal of Logic and Computation* 15, 559–574.
- Tarski, A., 1951, *A Decision Method for Elementary Algebra and Geometry*. Berkeley, Univ. California Press.

# On Field's Nominalization of Physical Theories

**Abstract.** In his book *Science Without Numbers* Harry Field argues that we can “nominalize” our physical theories, that is we can reformulate them in such a way that (1) the new version preserves the attractivity of the theory, and (2) the nominalized theory does not contain quantification over mathematical entities. I reconsider Field's nominalization procedure for a toy physical theory formulated in a first order language, in order to make a clear distinction between the following three steps: (1) the physical theory in terms of empirical observations; (2) the standard physical theory, which contains quantification over mathematical entities, as usual; (3) the nominalized version of the theory without any reference to mathematical entities. Having Field's nominalization procedure reconstructed, it will be clear that from a formalist point of view there is no difference between the original and the nominalized versions of the theory. It is because the only difference would come from the different “meanings” of the variables over which the quantification is running. The formalist philosophy of mathematics, however, denies that the variables have meanings at all. Finally, I consider further arguments for and against the indispensability of mathematical objects.

## 1 INTRODUCTION

One of the most important questions in the philosophy of mathematics is the ontological status of mathematical entities. In the late 1970s, Quine and Putnam suggested an argument for the existence of mathematical entities. The argument is based on the idea that mathematics is not only applicable but in fact *indispensable* for the empirical sciences; and if that is the case, then mathematical entities are as indispensable for our best ontological picture of the world as electrons and other physical entities, to the existence of which physicists are committed.

In its most explicit form the argument reads as follows:

### Indispensability argument

(P1) We ought to have ontological commitment to all and only the entities that are indispensable to our best scientific theories.

(P2) Mathematical entities are indispensable to our best scientific theories.

---

(C) We ought to have ontological commitment to mathematical entities.

The argument has attracted a great deal of attention. Many Platonists regard it as the best available argument for the existence of mathematical entities. The opponents of the argument object, first of all, to the first premise; while the second premise is considered uncontroversial (Colyvan 2004).

What does the first premise exactly mean? First of all, it definitely presupposes a kind of naturalism. For naturalism claims that we have to have ontological commitment to all and only the entities that exist according to our best scientific theories. According to Quine—see (Quine 1961), (Quine 1981)—if the language of the scientific theory quantifies over some entities which are, at the same time, indispensable, then we ought to have ontological commitment to those entities. It is therefore necessary to clarify the proper meaning of “indispensability”.

It would be quite straightforward to interpret dispensability as eliminability. An entity is eliminable from a theory if there is another theory which is empirically equivalent to the original one, but does not quantify over the entity in question. In this case, however, every non-observable entity would be dispensable, due to the well-known Craig theorem; see (Craig 1953). In other words, we would have to reject the existence of anything but the directly observable entities. In order to avoid such a radical conclusion, many suggest, we need to prescribe some further requirements for the new theory. These requirements are usually the following: clarity, simplicity, unificatory power, generality and fecundity ((Burgess 1983), (Maddy 2005)). These requirements altogether are called *attractivity*. So, an entity is dispensable if it is eliminable and the theory we obtain by its elimination remains attractive. In any event, the notion of attractivity is quite ambiguous; and it is hard to believe that the most fundamental ontological questions depend on such unclear and sociologically relative notions.

On the other hand, the second premise was considered as an evident one. Harry Field was the first who claimed that the second premise is, in fact, false; mathematical entities are not indispensable to our best scientific theories. Field adopted Quine’s linguistic criterion that a scientific theory asserts the existence of an entity by quantifying over the entity in question. He also accepted the attractivity requirements. But he showed that a physical theory can always be “nominalized”, by which he means that it can be reconstructed such that (1) the new theory does not contain quantification over mathematical entities, however, (2) remains attractive.



## 2 THE NOMINALIZATION PROCEDURE

In his *Science Without Numbers* (Field 1980) Field showed, as an example, how a fragment of Newtonian gravitational theory can be nominalized. In what follows, I will reconstruct Field's nominalization procedure in the case of an even simpler "toy" physical theory. My purpose is not only to demonstrate the nominalization steps on a perhaps more clear-cut example, but also to lay the emphasis on different points. First, within a physical theory, we will make a clear separation of the formal system and the semantics. Second, we will keep it clear that the equivalence of physical theories is understood as *empirical* equivalence (Fig. 1).

The nominalization procedure consists of the following three steps:

- (1) We have a body of physical facts, in terms of empirical observations.
- (2) We have the usual platonistic physical theory describing the observable phenomena in question—containing quantification over mathematical entities. The platonistic theory will consist of a formal system  $L$ , and a semantics  $S$ .
- (3) We construct a new theory which is capable of describing the same phenomena, but without quantification over mathematical entities. The nominalized theory will consist of a formal system  $L'$ , and a semantics  $S'$ . We will show the equivalence of these theories on the level of observable phenomena.

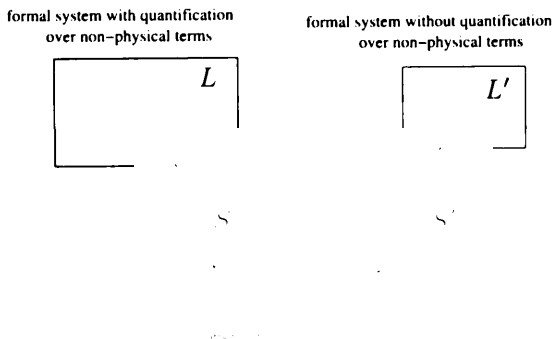


Figure 1. The platonistic physical theory and the nominalized physical theory should be equivalent on the empirical level.

### 3 THE “TOY” PHYSICAL THEORY

The “toy” physical theory is about a few empirically observable regularities related with the spatial relations of properties of the material points/molecules of a sheet of paper. Thus the only physical entities are the molecules of the paper and the only measuring equipment will be a scale-free ruler. We will examine—empirically—the following two properties of the molecules:

**Betweenness** We say that molecule  $\gamma$  is between molecules  $\alpha$  and  $\beta$  if whenever the ruler fits to  $\alpha$  and  $\beta$  then it also fits to  $\gamma$  and the mark on the ruler corresponding to  $\gamma$  falls between the marks corresponding to  $\alpha$  and  $\beta$ .

**Congruence** We will say that a pair of molecules  $\alpha, \beta$  is congruent to a pair of molecules  $\gamma, \delta$  if whenever we mark the ruler at  $\alpha, \beta$ , the same marks will also fit to  $\gamma, \delta$ .

With our scale-free ruler we can observe the following empirical facts about the molecules of the paper:

- (E1) If molecules  $\alpha$  and  $\beta$  are congruent to molecules  $\gamma$  and  $\delta$ , and  $\gamma$  and  $\delta$  are congruent to molecules  $\varepsilon$  and  $\zeta$ , then  $\alpha$  and  $\beta$  are congruent to  $\varepsilon$  and  $\zeta$ .
- (E2) If we consider three molecules fitting to the ruler, then there is exactly one that lies between the other two.

### 4 THE USUAL PLATONISTIC PHYSICAL THEORY OF THE PAPER

We present a “platonistic” physical theory  $(L, S)$  which describes the empirical facts of the paper. The formal system of the physical theory will be  $L = (\mathbb{R}^2, \Gamma, \Lambda)$ , where

$$\begin{aligned} \Gamma(a, b, c) \text{ is a relation between three points of } \mathbb{R}^2 \text{ (six real numbers):} \\ \Gamma(a, b, c) \iff \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2} + \sqrt{(c_1 - b_1)^2 + (c_2 - b_2)^2} \\ = \sqrt{(a_1 - c_1)^2 + (a_2 - c_2)^2} \end{aligned}$$

$$\begin{aligned} \Lambda(a, b, c, d) \text{ is a relation between four points of } \mathbb{R}^2 \text{ (eight real numbers):} \\ \Lambda(a, b, c, d) \iff \end{aligned}$$

$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2} = \sqrt{(c_1 - d_1)^2 + (c_2 - d_2)^2}$$

The semantics  $S$  is defined as follows: First, to every molecule we assign an element of  $\mathbb{R}^2$ :  $\alpha$  corresponds to  $a = (a_1, a_2) \in \mathbb{R}^2$ ,  $\beta$  to  $b = (b_1, b_2) \in \mathbb{R}^2$ , and so on. Second, relation  $\Gamma$  corresponds to the *Betweenness* and  $\Lambda$  corresponds to the *Congruence* of the molecules. All this representation is carefully constructed so that the physical theory  $(L, S)$ , that is, the formal system  $(\mathbb{R}^2, \Gamma, \Lambda)$  with the

above semantics provides a proper description of our empirical knowledge about the paper. It means that if  $\Gamma(a, b, c)$  is true for  $a, b, c \in \mathbb{R}^2$  then it is true for the corresponding molecules  $\alpha, \beta$  and  $\gamma$  that molecule  $\beta$  is between  $\alpha$  and  $\gamma$ . Similarly, if  $\Lambda(a, b, c, d)$  is true for  $a, b, c, d \in \mathbb{R}^2$  then it is true that molecules  $\alpha, \beta$  are congruent to molecules  $\gamma, \delta$ .

For example, empirical facts (E1) and (E2) are obviously represented in the theory  $(L, S)$ . Moreover,  $(L, S)$  has predictive power. For instance, in  $(\mathbb{R}^2, \Gamma, \Lambda)$  we can prove the following theorem (Fig. 2):

**Theorem 1.**

$$\begin{aligned} & \forall a \forall b \forall g \forall d \forall e \forall z \exists o \Gamma(a, d, b) \wedge \Gamma(b, e, g) \wedge \Gamma(g, z, a) \\ & \wedge \Lambda(a, d, d, b) \wedge \Lambda(b, e, e, g) \wedge \Lambda(g, z, z, a) \\ & \rightarrow \Gamma(a, o, e) \wedge \Gamma(b, o, z) \wedge \Gamma(g, o, d) \end{aligned}$$

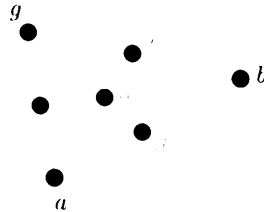


Figure 2. The centroid of a triangle always exists.

With the above semantics, we arrive at the following *hypothesis* about the molecules of the paper (Fig. 3):

**Hypothesis** If molecules  $\alpha, \beta, \gamma, \delta, \varepsilon$  and  $\zeta$  satisfy that  $\delta$  is between  $\alpha$  and  $\beta$ ,  $\varepsilon$  is between  $\beta$  and  $\gamma$ , and  $\zeta$  is between  $\gamma$  and  $\alpha$ , furthermore,  $\alpha, \delta$  are congruent to  $\delta, \beta$ , and  $\beta, \varepsilon$  are congruent to  $\varepsilon, \gamma$ , and  $\gamma, \zeta$  are congruent to  $\zeta, \alpha$ , then we can always find a molecule  $\omega$  such that it is between  $\alpha$  and  $\varepsilon$ , and it is between  $\beta$  and  $\zeta$  and it is between  $\gamma$  and  $\delta$ .

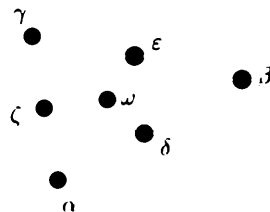


Figure 3. According to the semantics of the theory, Theorem 1. (Fig. 2) is a statement about the molecules of the paper.

This hypothesis can be verified empirically by means of the scale-free ruler; and we will find that the hypothesis is true.

In Field's terminology,  $(L, S)$  is a *platonistic* theory: It contains quantification over mathematical entities, namely, over real numbers, since  $\Gamma$  and  $\Lambda$  are relations between real numbers.

## 5 THE NOMINALIZED THEORY

Now we will construct another physical theory which can equally well describe the same observable phenomena but without quantification over mathematical entities. This will consist of another formal system  $L'$  with another semantics  $S'$ .

The formal language  $L'$  will be a first order formal system.  $L'$  will contain individuum variables  $A, B, C, \dots$  and a three-argument predicate symbol *Bet* and a four-argument predicate symbol *Cong*. Beyond the logical axioms of PC(=) (predicate calculus with identity), we will need the following "physical" axioms:<sup>1</sup>

$$\mathbf{T1:} \quad \forall A \forall B \text{ Cong}(A, B, B, A)$$

$$\mathbf{T2:} \quad \forall A \forall B \forall C \text{ Cong}(A, B, C, C) \rightarrow A = B$$

$$\mathbf{T3:} \quad \forall A \forall B \forall C \forall D \forall E \forall F \text{ Cong}(A, B, C, D) \wedge \text{Cong}(C, D, E, F) \\ \rightarrow \text{Cong}(A, B, E, F)$$

$$\mathbf{T4:} \quad \forall A \forall B \text{ Bet}(A, B, A) \rightarrow A = B$$

$$\mathbf{T5:} \quad \forall A \forall B \forall C \forall D \forall E \text{ Bet}(A, D, C) \wedge \text{Bet}(B, E, C) \\ \rightarrow \exists F (\text{Bet}(D, F, B) \wedge \text{Bet}(E, F, A))$$

$$\mathbf{T6:} \quad \exists E \forall A \forall B \phi(A) \wedge \psi(B) \rightarrow \text{Bet}(E, A, B) \\ \rightarrow \exists F \forall A \forall B \phi(A) \wedge \psi(B) \rightarrow \text{Bet}(A, F, B)$$

where  $\phi$  and  $\psi$  are two arbitrary formulas of the language, containing no free instances either  $E$  or  $F$ . Let there also be no free instances of  $A$  in  $\psi(B)$  or of  $B$  in  $\phi(A)$ .

$$\mathbf{T7:} \quad \exists A \exists B \exists C \neg \text{Bet}(A, B, C) \wedge \neg \text{Bet}(B, C, A) \wedge \neg \text{Bet}(C, A, B)$$

$$\mathbf{T8:} \quad \forall A \forall B \forall C \forall D \forall E \text{ Cong}(A, D, A, E) \wedge \text{Cong}(B, D, B, E) \\ \wedge \text{Cong}(C, D, C, E) \wedge \neg D = E \\ \rightarrow \text{Bet}(A, B, C) \vee \text{Bet}(B, C, A) \vee \text{Bet}(C, A, B)$$

$$\mathbf{T9:} \quad \forall A \forall B \forall C \forall D \forall E \text{ Bet}(A, D, E) \wedge \text{Bet}(B, D, C) \wedge \neg A = D \\ \rightarrow \exists F \exists G \text{ Bet}(A, B, F) \wedge \text{Bet}(A, C, G) \wedge \text{Bet}(G, E, F)$$

$$\mathbf{T10:} \quad \forall A \forall B \forall C \forall D \forall E \forall F \forall G \forall H \neg A = B \wedge \text{Bet}(A, B, C) \wedge \text{Bet}(E, F, G) \\ \wedge \text{Cong}(A, B, E, F) \wedge \text{Cong}(B, C, F, G) \wedge \text{Cong}(A, D, E, H) \\ \wedge \text{Cong}(B, D, F, H) \rightarrow \text{Cong}(C, D, G, H)$$

$$\mathbf{T11:} \quad \forall A \forall B \forall C \forall D \exists E \text{ Bet}(D, A, E) \wedge \text{Cong}(A, E, B, C)$$

<sup>1</sup>The reader may recognize that these are nothing but the well known Tarski–Givant (1999) axioms of Euclidean geometry. But it must be emphasized that this fact is irrelevant.

The  $S'$  semantics of the theory is defined as follows. The individual variables  $A, B, C, \dots$  will refer to the molecules of the paper. The predicate symbol  $Bet$  corresponds to the *Betweenness*, and the predicate symbol  $Cong$  corresponds to the *Congruence*.

The physical theory  $(L', S')$  with the above semantics provides a proper description of our empirical knowledge about the paper. For example, empirical facts (E1) and (E2) are obviously represented by theorems in  $(L', S')$ . This theory equally well describes our empirical knowledge about the paper. It also has the same predictive power. For instance, in  $L'$  we can prove the following theorem Fig. 4:

**Theorem 1'**

$$\begin{aligned} & \forall A \forall B \forall G \forall D \forall E \forall Z \exists O \text{ Bet}(A, D, B) \\ & \wedge \text{Bet}(B, E, G) \wedge \text{Bet}(G, Z, A) \wedge \text{Cong}(A, D, D, B) \\ & \wedge \text{Cong}(B, E, E, G) \wedge \text{Cong}(G, Z, Z, A) \\ & \rightarrow \text{Bet}(A, O, E) \wedge \text{Bet}(B, O, Z) \wedge \text{Bet}(G, O, D) \end{aligned}$$

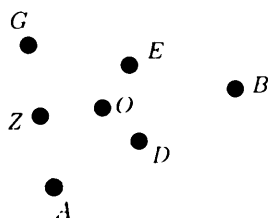


Figure 4. The centroid of a triangle always exists.

With the above semantics, this leads us to the following hypothesis about the molecules of the paper (Fig. 5):

**Hypothesis** If molecules  $\alpha, \beta, \gamma, \delta, \varepsilon$  and  $\zeta$  satisfy that  $\delta$  is between  $\alpha$  and  $\beta$ ,  $\varepsilon$  is between  $\beta$  and  $\gamma$ , and  $\zeta$  is between  $\gamma$  and  $\alpha$ , furthermore,  $\alpha, \delta$  are congruent to  $\delta, \beta$ , and  $\beta, \varepsilon$  are congruent to  $\varepsilon, \gamma$ , and  $\gamma, \zeta$  are congruent to  $\zeta, \alpha$ , then we can always find a molecule  $\omega$  such that it is between  $\alpha$  and  $\varepsilon$ , and it is between  $\beta$  and  $\zeta$  and it is between  $\gamma$  and  $\delta$ .

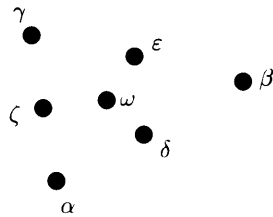


Figure 5. According to the semantics of the theory, 'Theorem 1' (Fig. 4) is a statement about the molecules of the paper.

This hypothesis can be verified empirically by means of the scale-free ruler; and we will find that the hypothesis is true.

## 6 CONCLUDING REMARKS

As we can see, quantification over the mathematical entities can be eliminated from a physical theory. For example, in this “toy” physical theory we eliminated quantification over the points of  $\mathbb{R}^2$  and quantification over the real numbers. This does not mean, however, that we have really purified our physical theory from Platonic objects: Although we eliminated *quantification* over mathematical entities, we did not eliminate the *mathematical structures* themselves. We still need the structure defined by axioms  $\Gamma 1$ – $\Gamma 11$ . It thus seems unavoidable to draw the conclusion that we ought to have ontological commitment to formal systems as abstract entities; and this is sufficient for the structuralist version of Platonism (Shapiro 1997).

It must be noted that both the Quine–Putnam argument and Field’s criticism are based on the tacit assumption that the terms and statements of mathematics have *meanings*, and the only question is the ontological status of the entities that mathematics refers to. According to the formalist philosophy of mathematics, however, this assumption is unacceptable, *ab ovo*.

As we have seen in our example, both the *platonistic* and the *nominalized* versions of the physical theory have the same structure: a *meaningless formal system* + a *partial semantics* pointing only to physical, moreover, observable things. From this point of view, it does not matter whether or not the formal system in question contains quantification over certain variables. Formal systems are obviously indispensable from both platonistic and nominalized physical theories, in spite of the fact that they are meaningless. The only question is: What is the ontological status of formal systems? And still, one can answer this question from a structuralist–Platonist position, drawing support from the Quine–Putnam indispensability argument. Or, one can consider an entirely different account

for formal systems, which completely intact from the indispensability argument (see for example (Szabó 2003) for a physicalist ontology of formal systems).\*

#### REFERENCES

- Burgess, John, 1983, Why I am not a nominalist. *Notre Dame J. Formal Logic* 24(1), 93-105.
- Colyvan, Mark, 2004, Indispensability arguments in the philosophy of mathematics. *The Stanford Encyclopedia of Philosophy* (Fall 2004 Edition)
- Craig, William, 1953, On axiomatizability within a system. *The Journal Of Symbolic Logic* 18, 30-32.
- Field, Harry H., *Science Without Numbers*. Basil Blackwell, Oxford.
- Maddy, Penelope, 2005, Three forms of naturalism. in Shapiro, S. (ed.), *The Oxford Handbook of Philosophy of Mathematics and Logic*. Oxford University Press.
- Quine, Willard V. O., 1961, On what there is. In *From a Logical Point of View*, 2nd ed. Harvard University Press.
- Quine, Willard V. O., Things and their places in theories. In *Theories and Things*. The Belknap Press of Harvard University Press.
- Shapiro, Stewart, 1997, *Philosophy of Mathematics: Structure and Ontology*. Oxford University Press, Oxford.
- Szabó, László E., 2003, Formal systems as physical objects: A physicalist account of mathematical truth, *International Studies in the Philosophy of Science* 17, 117-125.
- Tarski, Alfred and Givant, Steven, 1999, Tarski's system of geometry. *Bulletin of Symbolic Logic* 5, 175-214.

\*I would like to thank László E. Szabó for the conversations and insights which led up to this paper. I would also like to thank Réka Bence and Kristóf Szabó for their help.

# Contributors

- HAJNAL ANDRÉKA • Rényi Institute of Mathematics, Budapest  
ANNA BROŽEK • Department of Logical Semiotics, Warsaw University  
FERENC CSABA • Institute of Philosophy, Eötvös Loránd University, Budapest  
GÁBOR FORRAI • Department of Philosophy, University of Miskolc  
ZOLTÁN GENDLER SZABÓ • Department of Philosophy, Yale University, New Haven  
ROBIN HIRSCH • Department of Computer Science, University College London  
LÁSZLÓ KÁLMÁN • Department of Theoretical Linguistics, Eötvös Loránd University, Budapest / Hungarian Academy of Sciences  
EDWARD KANTERIAN • Trinity College / Jesus College, Oxford  
GYULA KLÍMA • Department of Philosophy, Fordham University, New York  
ANDRÁS KORNAI • Media Research Center, Budapest University of Technology and Economics  
JUDIT X. MADARÁSZ • Rényi Institute of Mathematics, Budapest  
ANDRÁS MÁTÉ • Department of Logic, Eötvös Loránd University, Budapest  
PÉTER MEKIS • Department of Logic, Eötvös Loránd University, Budapest  
TAMÁS MIHÁLYDEÁK • Department of Computer Science, University of Debrecen  
NENAD MISCEVIC • Department of Philosophy, Central European University, Budapest / Department of Philosophy, University of Maribor  
ISTVÁN NÉMETHI • Rényi Institute of Mathematics, Budapest  
EDI PAVLOVIĆ • Faculty of Philosophy, University of Rijeka  
JIRÍ RAČLAVSKÝ • Department of Philosophy, Masaryk University, Brno  
MÁTÉ SZABÓ • Department of Philosophy and History of Science, Budapest University of Technology and Economics  
ANNA SZABOLCSI • Department of Linguistics, New York University  
GERGELY SZÉKELY • Rényi Institute of Mathematics, Budapest  
MÁRTA UJVÁRI • Department of Philosophy, Corvinus University, Budapest  
ZSÓFIA ZVOLENSZKY • Department of Logic, Eötvös Loránd University, Budapest



A MAGYAR FILOZÓFIAI SZEMLE JÖVŐ ÉVI TEMATIKUS SZÁMAI  
ÉS A KÉZIRATOK LEADÁSÁNAK HATÁRIDEJE:

*Irracionalitás* (2011/1) január 31.

*Hatalom* (2011/2) május 15.

*Tapasztalat* (2011/3) szeptember 15.

## SZERZŐINK FIGYELMÉBE

A *Magyar Filozófiai Szemle* referált folyóirat, amely csak eredeti kéziratokat közöl, és minden beérkezett tanulmányt két független, anonim bírálónak ad ki. A kéziratokat elektronikus formában várjuk a [fiilszemle@webmail.phil-inst.hu](mailto:fiilszemle@webmail.phil-inst.hu) címre. A teljes példány mellé egy másik, név és önhivatkozások nélküli változatot is kérünk a referensek részére.

A kéziratokat Times New Roman betűtípusban, „rtf” formátumban, körülbelül egyívnyi (40 000 leütés) terjedelemben, csupán a legszükségesebb formázásokkal (például kurzíválás) kérjük. A főszövegben a 4-5 sornál terjedelmesebb idézetek a főszövegtől egy-egy sorral elválasztva, behúzva szerepeljenek.

Tanulmányok esetében a hivatkozott szakirodalmat a szövegbe illesztve kérjük feltüntetni: (Kuhn 1999a. 99), vagy lábjegyzetben: Putnam 1975b. 237. Szükség esetén megadható az eredeti megjelenés dátuma és a kötetszám is (Descartes 1644/1996. 8/2:72). Használhatók a széles körben elterjedt rövidítések (AT VIII-2. 72), és klasszikus szövegek esetében a bevett utalási eljárások (Platón, *Phaidrosz* XXIV 245c–d; Mt 9,13). A rövidítések feloldását kérjük az íráskor végén, az irodalomjegyzékben megadni.

Az irodalomjegyzéket a következő példákknak megfelelően kérjük összeállítani:

Önálló kötet:

Locke, John 1690/2003. *Értekezés az emberi értelemről*. Ford. Vassányi Miklós – Csordás Dávid. Budapest, Osiris.

Tanulmány gyűjteményes kötetben:

Henke, Christopher R. – Thomas F. Gieryn 2008. Sites of Scientific Practice: The Enduring Importance of Place. In Edward J. Hackett – Olga Amsterdamska – Michael Lynch – Judy Wojcman (szerk.) *The Handbook of Science and Technology Studies*. Cambridge/MA, The MIT Press. 354–376.

Periodikumban megjelent cikk, tanulmány:

Schnädelbach, Herbert 1993. A filozófia és az embertudományok. Ford. Mezei György Iván. *Magyar Filozófiai Szemle*, 37/2. 1015–1031.

A kéziratokhoz rövid bemutatkozást és angol nyelvű összefoglalót kérünk mellékelni.

---

# HUNGARIAN PHILOSOPHICAL REVIEW

2010/4 (54. évfolyam)

---

## MAGYAR FILOZÓFIAI SZEMLE

This special issue of the *Hungarian Philosophical Review* is dedicated to the memory of our friend, colleague and teacher Imre Ruzsa (1921–2008). Ruzsa was the father of modern philosophical logic in Hungary, the founder and the first chair of the Department of Logic at Eötvös University. His professional interests centered around modal logic, intensional logic, modeling natural language in systems of intensional logic, and the foundations of logic and mathematics. He always thought of his generalization of A. N. Prior's concept of semantic value gaps to quantified, intensional and type-theoretic systems as his most important contribution to logic. He was the author of three books in English (*Modal Logic with Descriptions*, The Hague, 1982; *Intensional Logic Revisited*, Budapest, 1991; *Introduction to Metalogic*, Budapest, 1993), several monographs and textbooks in Hungarian, and many articles in leading logic journals.

The variety of the papers included in this volume represents the range of topics that Ruzsa's research covered: philosophical logic, formal semantics, the philosophy of mathematics, and foundational studies. The papers were all given at the Imre Ruzsa Memorial Conference "Logic, Language, Mathematics", held at the Philosophy Institute of Eötvös University in Budapest on September 17–19, 2009, as part of the annual conference series "Language, Understanding, Interpretation".

ISSN 0025-0090



1200 Ft

**GONDOLAT**

**nka**  
Nemzeti Kulturális Alap