



INDETERMINACY

AND

PERSONAL IDENTITY

Matthew Harvey

Discipline of Philosophy
The University of Adelaide

Submitted for the degree of Doctor of Philosophy

January 2006

Dedicated to the memory of Gladys Hilda Harvey

Contents

Abstract	v
Statement	vii
Acknowledgments	viii
<i>Introduction</i>	<i>1</i>
<i>1 Personal identity and indeterminacy</i>	<i>3</i>
1.1 Introduction	3
1.2 Some terminology	4
1.3 Indeterminate statements of non-personal diachronic identity	15
1.4 The Combined Spectrum Argument	19
1.5 Goodenough's critique of Parfit	32
1.6 Williams's Torture Case	39
1.7 A semantic framework	48
1.8 Conclusion	58
<i>2 Vagueness</i>	<i>59</i>
2.1 Introduction	59
2.2 Sorites paradoxes and theories of vagueness	63
2.3 Williamson's argument for Bivalentism	79
2.4 The coherence of Anti-Determinist Classical Bivalentism	87
2.5 Some challenges for Anti-Determinist Classical Bivalentism	97
2.6 The Determinist's explanatory burden	104
2.7 The Anti-Determinist Classical Bivalentist's explanatory burden	114
2.8 The Omniscient Speakers Argument and higher-order vagueness	127
2.9 Conclusions	137

3	<i>The Simple View</i>	139
3.1	Introduction	139
3.2	Some initial arguments	143
3.3	Williams's twin thought experiments	148
3.4	A broader argument for the Simple View	161
3.5	Other responses to Williams	172
3.6	Conclusion	174
3.7	Some further remarks	175
4	<i>Fission</i>	178
4.1	Introduction	178
4.2	The Fission Case	179
4.3	Responses to the case	182
4.4	Garrett's critique of the indeterministic response to Fission	192
4.5	Practical considerations	211
4.6	Conclusion	215
	<i>Conclusion</i>	216
	<i>Bibliography</i>	217

Abstract

Chapter 1

A statement '*S*' is *indeterminate* if and only if the facts do not determine that *S*, and the facts do not determine that not-*S*. Statements of non-personal diachronic identity can be indeterminate, unless the epistemic view of vagueness is correct. If statements of non-personal diachronic identity can be indeterminate, then statements of diachronic personal identity (DPI) can be indeterminate, unless either (a) the Simple View of DPI both is true and is an effective way of resisting Parfit's (1984) "Combined Spectrum Argument", or (b) the first-person perspective that appears distinctive of personal identity makes DPI an appropriately special kind of diachronic identity. There is no good argument for (b). Also, Goodenough (1996) is mistaken in his critique of Parfit's (1984) argumentative deployment of his "Spectra".

Chapter 2

Some ways of resisting Williamson's (1994) *argument against supposed counterexamples to Bivalentism* are better than others. But even Classical Bivalentism does not entail Williamson's epistemic view of vagueness: Indeterminist Classical Bivalentism (ICB) is coherent. To be *plausible* ICB must be compatible with a non-circular, non-epistemic explanation of *indeterminacy*. The Determinist also bears a certain explanatory burden. But while Determinists' attempts to discharge their burden have been unconvincing, the defender of ICB can explain indeterminacy satisfactorily. Since ICB is plausible, the rejection of Classical Bivalentism is unmotivated. I also discuss Williamson's (1994) "Omniscient Speakers Argument".

Chapter 3

We should disbelieve the implausible Simple View of DPI pending a good argument in its favour. Williams (1970) contains no such argument. Other arguments for the Simple View are also seen to fail. Defeating these arguments brings to light the important distinction between variation with respect to different possible situations,

and variation with respect to different presentations or descriptions of a single situation.

Chapter 4

I discuss the question of “who is who” in Wiggins’s (1967) “Fission Case”. There are good reasons to hold that this case gives rise to indeterminate personal identity statements. I defend this view from criticisms by Garrett (1998).

Statement of Originality

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

10 May 2006

Acknowledgments

Thanks, especially, to Garrett Cullity, for unwavering support, encouragement and critical feedback of the highest quality. Thanks also to Chris Mortensen, for much encouragement and assistance, as well as for introducing me to so much interesting philosophy during my undergraduate years. Thanks to Nicole Vincent and Elizabeth Schier, for some helpful conversations, and for Nicole's computer-related assistance; to Anne Harvey for conversations pertaining to Chapter 4; to Andrew Turner for help regarding perdurantism and endurantism; and to Philip Gerrans, for his feedback in Review Panel meetings.

INTRODUCTION

What follows concerns two topics: *vagueness*; and *diachronic personal identity*.

Here are some paradigm *vague predicates*: ‘is bald’; ‘is a heap’; ‘is tall’. These predicates admit of borderline cases. For example, some people are neither clearly tall, nor clearly not tall.

A statement of *diachronic personal identity*, put roughly, is a statement that says that a person existing at one time is one and the same individual as a person existing at some other time.

The primary aims in each of the chapters are as follows; some secondary aims will become apparent in the chapters themselves.

Chapter 1 makes an initial case for the claim that statements of diachronic personal identity can be indeterminate, the facts neither determining that they hold, nor determining that they fail to hold. Parfit’s (1984) Combined Spectrum Argument for this claim is expounded, and various threats to this argument are defused.

I deal with the general topic of vagueness in Chapter 2. Here, I attack the view that the apparent unclarity surrounding the borderline cases of vague predicates is in fact an epistemic phenomenon: our ignorance. This counterintuitive view has received much attention in recent years thanks to challenging arguments in its favour by Timothy Williamson (esp. 1994), on whose work I shall focus, and Roy Sorensen (1988, 2001a).

To confront the epistemic view properly, one must confront also the challenge of explaining the notion of non-epistemic *indeterminacy*. Chapter 2, besides being an attack on the epistemic view, is intended as a defence of a classical and bivalent, yet non-epistemic treatment of vagueness. In the course of this defence, I offer a non-epistemic, non-circular explication of the notion of indeterminacy.

Chapter 2 operates also as a defence, from the general threat posed by the epistemic view, of the specific claim, propounded in Chapter 1, that there can be genuinely indeterminate statements of diachronic personal identity.

Chapter 3 addresses arguments for the so-called Simple View of personal identity, which view, if true, would undermine the argument of Chapter 1. Though the Simple View is (rightly) rejected by most philosophers nowadays, examining certain arguments for it remains an important exercise. It brings to light the importance of the distinction – touched upon also in Chapter 2 – between (a) variation with respect to different possible situations, and (b) variation with respect to how a given situation might be represented or described.

In Chapter 4 I consider the much-discussed *Fission Case*. In this hypothetical case – put roughly and somewhat question-beggingly – one person, *a*, splits, symmetrically, into two persons, *b* and *c*. I defend a controversial view, according to which *a* determinately survives fission, but it is indeterminate whether $a=b$, and it is indeterminate whether $a=c$. Some of the lessons of Chapter 2 are applied in defending this view.

I leave many questions regarding personal identity untouched. For example, I do not address the question of whether the continued existence of a person consists in psychological continuity, physical continuity, or some combination of the two.

I do not address at length the controversial issue of the normative significance, or insignificance, of the diachronic personal identity relation; although there are points at which I touch upon this matter.

Chapter 1

PERSONAL IDENTITY AND INDETERMINACY

1.1 INTRODUCTION

The aim of this chapter is to make an initial case for the following claim.

(Personal Indeterminism) There can be indeterminate statements of diachronic personal identity.

In Section 1.2, I will define ‘statement of diachronic personal identity’ and explain what it would be for one of these statements to be ‘indeterminate’. In Section 1.3, I will argue that statements of *non*-personal diachronic identity can be indeterminate. In Section 1.4, I will describe Parfit’s (1984: 236-243) “Combined Spectrum Argument” for Personal Indeterminism, and will address some initial objections to that argument by Madell (1985). In Section 1.5, I will consider a critique of Parfit’s reasoning by Goodenough (1996), and will explain how this critique is seriously flawed. In Section 1.6 I will consider whether there is reason to think that statements of diachronic personal identity (DPI) are immune from indeterminacy on account of the first-person perspective associated with DPI, but not with other forms of diachronic identity. I will pay particular attention here to arguments by Williams (1970) and Noonan (1989). I will conclude by agreeing with Noonan that the first-person perspective does not make DPI appropriately special. In Section 1.7 I will distinguish three accounts of the source of the indeterminacy of indeterminate diachronic identity statements, and will give some reasons for preferring a particular one of these accounts. In Section 1.8, I will summarise the conclusions of the present chapter. I will conclude that Personal Indeterminism is true *unless* either (a) the epistemic view of vagueness is correct, or (b) a certain “Simple View” about personal identity both is true, and is an effective

way of blocking the Combined Spectrum Argument. I will address (a) in Chapter 2, and (b) in Chapter 3.

1.2 SOME TERMINOLOGY

1.2.1

What is a statement of diachronic personal identity, and what would it be for such a statement to be indeterminate? My investigations need not await precise definitions of these terms; but with some discussion, and by offering examples, it should be possible to convey sufficiently clearly what I mean by these terms.

In particular, a precise definition of ‘indeterminate’ would be impossible without presupposing answers to some of the questions about the concept of indeterminacy that it will be my goal to investigate.

1.2.2 ‘Statement’ and schemas

First, what is meant by ‘statement’?

I will use the term ‘statement’ to mean ‘unambiguous strongly indicative utterance type’.

I use the term ‘strongly indicative utterance’ to refer to all and only those utterances which, in the words of Timothy Williamson (1994: 187), ‘say that something is the case’.

As Williamson points out, not all utterances of declarative sentences ‘say that something is the case’:

If a teacher pronounces ‘He was there then’ as a sample sentence of English, leaving ‘he’, ‘there’ and ‘then’ undetermined in reference, nothing has been said to be the case ... (1994: 187)

Williamson (187) also points out that an utterance can say that something is the case even if it is uttered other than in order to assert that something is the case. A speaker might, for example, *suppose* that something is the case. In making this supposition, the speaker makes an utterance that says that something (*viz.* what the speaker is

supposing to be the case) is the case, and that utterance is thus still a strongly indicative utterance.

Statements are *utterance types*, not individual utterances. However, two individual utterances are of *distinct* types (for the purposes of how I will be using ‘statement’), *i.e.* are tokens of two distinct statements, even if the only respect in which they differ is that the contexts in which they are uttered are relevantly different, *i.e.* are different in a way that effects a difference in meaning between the two individual utterances.

The following two individual printed utterances both belong to the same utterance type.

ONE PLUS ONE EQUALS TWO.

ONE PLUS ONE EQUALS TWO.

Even though each of the above two individual utterances occurs on a separate line, this does not effect any difference in meaning between them. However, the following two individual printed utterances belong to two distinct utterance types – and so are tokens of two distinct statements.

THE SENTENCE BELOW IS PRINTED IN BLACK.

THE SENTENCE BELOW IS PRINTED IN BLACK.

The question of whether or when such differences as differences in pragmatic factors, emotional overtones, or other similarly subtle factors, count as differences in meaning, is not a relevant question for my purposes.

Statements are by definition *unambiguous*. Each statement has a distinct meaning. An utterance is ambiguous if (given its context), it might have multiple distinct meanings, for example (if uttered in certain unusual contexts): ‘I drew some curtains.’ To an ambiguous utterance there does not correspond any statement. But ambiguity is not the same as vagueness. I will discuss vagueness later; but for now, simply note that *a statement may be vague*. Thus ‘Pope Benedict XVI is tall’ is unambiguous – it has a distinct meaning – even though it is vague. It is vague because

the predicate 'is tall' is vague. It is a statement because it says that something is the case: it says that Pope Benedict XVI is tall. However, the predicate 'is tall' is not ambiguous, and neither is 'Pope Benedict XVI'; and neither is this statement syntactically ambiguous.

Each statement has its unique context. I shall assume that, when explicit mention of context is not made, a certain *default context* accompanies all citings of linguistic expressions in philosophical discussions, in case the intent is thereby to refer to a statement, such that this default context is one and the same default context for all such citings, and such that linguistic items occurring in this context do not, in virtue of their occurrence here, carry meanings that are in any way abnormal or distorted in relation to the meaning(s) they usually do carry in ordinary conversation. (This does not mean that I cannot discuss statements in other contexts on occasion, by making explicit that I intend this.)

Statements can be true or false. However my definition of 'statement' leaves open the possibility that some statements are neither true nor false.

A *predication* is a statement resulting from the application of a predicate to one or more singular terms. For example, the application of the predicate 'is hot' to the singular term 'the sun', results in the following predication: 'The sun is hot.'

There may be occasions when the only way to make sense of one of my passages is as a passage-schema – in effect, an abbreviation for an infinite series of all and only the passages of the given form – even when I make no explicit mention of this fact. I try to make explicit mention of the schematic nature of certain passages when the passages are particularly important, such as in the definition of 'statement of diachronic personal identity' below. But to make explicit such mention all the time would involve much clutter. I hope, then, that any loss of rigour due to failure always to bind explicitly such schematic passages is compensated for by the reduction in clutter.

1.2.3 'Statement of diachronic personal identity'

I now define 'statement of diachronic personal identity' ('SDPI') as follows.

Every instance of the following schema is now stipulated to be analytically true, where '*P*', '*Q*', '*t*' and '*t**' are replaceable by inscriptions of singular terms.

(1AA) If P is a person, Q is any object (persons count among objects), t is a time and t^* is another time, P exists¹ at t and Q exists at t^* , the statement ' P is numerically identical² to Q , and P exists at t and Q exists at t^* ' is a statement of diachronic personal identity ("SDPI"), which statement I refer to using the expression ' $\wedge P, Q, t, t^*$ '.

Furthermore, any statement that is synonymous with $\wedge P, Q, t, t^*$ is a statement of diachronic personal identity.

The specification ' P exists at t and Q exists at t^* ' might merely be *implicit* in some such statements.

Furthermore, I use the term ' $\langle\langle P, Q, t, t^* \rangle\rangle$ ' to refer to an arbitrary statement among all of these statements.

An analogous definition may be given for the broader concept of a *statement of (personal or non-personal) diachronic identity*, by omitting, from the above definition, the stipulation that P is a person.

Note that it is not stipulated in 1AA "that Q is a person"; it is only stipulated "that P is a person". Some people think there might be an entity which at one time is a person, but which at some other time is not a person. Arguably I was once a foetus, and arguably foetuses are not persons. 1AA includes within its scope statements identifying a person existing at one time with a non-person existing at another time; and thus one might pursue, within one's investigations regarding what kinds of statements of diachronic personal identity can be true, the question of whether persons

¹ I often treat 'exists' as of no specific tense.

² To say that an object A is numerically identical to an object B is to say that A is one and the same object as B. Numeric identity is to be distinguished from qualitative identity. To say that A is qualitatively identical to B is to say that A and B are exactly alike – that they have all the same qualities. There might be two objects that are exactly alike. But if object A is numerically identical to object B, there are not two objects, but one. Unless otherwise specified, I always mean by 'identity' and its cognates to express numerical identity.

An unorthodox theory defended by Sider (*e.g.* 1996) analyses diachronic identity (or apparent identity) as a relation holding between (numerically) distinct temporal object-stages which are temporal "counterparts", rather than as the strict numerical identity relation borne either by a perduring aggregate of temporal stages to itself, or by an enduring object to itself. I will ignore this view, though I have nothing against it, in order mainly to avoid wordy provisos. I strongly suspect that Sider's brand of four-dimensionalism would give him no *particular* reason to reject a suitably reworded version of (at least) almost all of my assertions; but I do not propose to argue here that my suspecting this is justified.

are *essentially* persons. I myself will not be pursuing this question. (See Olson 2002 §2.)

To help clarify further what is meant by ‘statement of diachronic personal identity’, here is an example of an SDPI.

- (1A) The person in my study now is the person who was in my study yesterday.

The person in my study now is a person, and the person who was in my study yesterday is an object, now is a time, yesterday is another time, the person in my study now exists now, and the person who was in my study yesterday exists yesterday, and ‘is’ in 1A has the sense of numerical identity. So 1A is an SDPI – with the specification, *e.g.*, that the person who was in my study yesterday exists yesterday being implicit in this SDPI, in virtue of the ‘yesterday’ in the relevant definite description.

Given an SDPI $\langle P, Q, t, t^* \rangle$, in accordance with definition 1AA, it may even be that the specification that P exists at t and Q exists at t^* , which specification is required to be explicit or implicit in $\langle P, Q, t, t^* \rangle$, is effected *via* an explicit stipulation, made *outside* the statement $\langle P, Q, t, t^* \rangle$ itself, that the singular terms ‘ P ’ and ‘ Q ’ shall refer, respectively, to some individuals x and y that are specified in that stipulation by reference to their existing at times t and t^* respectively. *E.g.*, suppose I stipulate that ‘ J ’ shall refer to the person who is Prime Minister of New Zealand as at sunrise tomorrow, and that ‘ K ’ shall refer to whoever is Prime Minister of New Zealand as at sunset tomorrow. If I then assert ‘ J is K ’, I have asserted an SDPI in accordance with definition 1AA, since an implicit specification, for some distinct times t and t^* , that J exists at t and K exists at t^* , has been carried “into” my utterance of ‘ J is K ’ *via* my previous time-referring stipulations regarding to which persons ‘ J ’ and ‘ K ’ shall refer.

This may seem obvious, but it is worth emphasising that 1A is an SDPI regardless of its truth-value: 1A is still a *statement* of diachronic personal identity regardless of whether the person in the room now really *is* the same individual as the person in the room yesterday. An SDPI may be true; or it may be false. Given definition 1AA, given an SDPI $\langle P, Q, t, t^* \rangle$, the fact that it *is* an SDPI analytically entails that P exists at t and that Q exists at t^* . Thus, $\langle P, Q, t, t^* \rangle$ is false *only* on

condition that P is not numerically identical to Q . The question now is, might an SDPI be indeterminate?

1.2.4 'Indeterminate statement'

In this Subsection I aim to exhibit some of the constraints characterising the notion of *indeterminacy*, and related notions, without purporting to provide any reductive definition.

In particular, I hereby assert every instance of the following schema (*), such that: for each indexical-free statement s whose context is the default context, there is an instance of * which instance is the passage derived from * by substituting, for every occurrence of ' S ', an inscription of s ; and such that nothing else is an instance of *.

- (*) The statement ' S ' is *indeterminate* if, and only if, the facts (both about the world apart from language, and about the meaning and use of the words, expressions *etc.* involved in ' S ') both fail to determine that S , and fail to determine that it is not the case that S .

One of things that I have said, for example, in one of the infinitely many passages just effected, is that the statement 'The moon is made of solid gold' is indeterminate if, and only if, the facts (both about the meaning and use of the words 'moon', 'gold' *etc.*, and about the physical constitution of the moon itself) both fail to determine that the moon is made of solid gold, and fail to determine that the moon is not made of solid gold.

* is intended to characterise *directly* the notion of indeterminacy only for statements of the default context that are free of indexicals. But the scope of this notion extends to *all* statements. *I.e.*, statements containing indexicals, or not of the default context, can also meaningfully (whether or not truly) be said to be indeterminate, or said not to be indeterminate. And I assume analogous schemata could be provided for statements of all contexts and containing any indexicals.

It would have been wrong for me to have said that the statement uttered by the Pope in *his* utterance of 'I am bald' is indeterminate if and only if the facts both fail to determine that I am bald and fail to determine that I am not bald. *That* statement by the Pope contains an indexical ('I'); and so no inscription of it is an instance of *.

Nevertheless, one of the infinitely many passages effected above tells us that, *if* the facts fail to determine either that the Pope is bald or that he is not bald, *then* ‘the Pope is bald’ (in the default context) is an indeterminate statement. By comparison with that statement, one can see that, assuming the Pope refers to himself in uttering ‘I’, *his* utterance of ‘I am bald’ would be an utterance of an indeterminate statement if and only if the facts fail to determine that *the Pope* is bald, and fail to determine that *the Pope* is not bald.

It is a mark of indeterminacy that, even if we knew all the facts relevant to an indeterminate statement, we should still be unable to say confidently either that the statement is true, or that it is false – despite being perfectly familiar with the statement’s meaning. If ‘*S*’ is indeterminate, then we, in being confronted with this statement, feel unsure or unclear about what to say in response to the question ‘Is it the case that *S*?’ if what is demanded of us is either a ‘Yes’ or a ‘No’. This felt “uncertainty” is not due to ignorance: simply coming to know all the facts would not remove our uncertainty. Thus, indeterminacy is not an epistemic phenomenon. There are those, *e.g.* Timothy Williamson (1994) and James Cargile (1969), who hold that *vagueness* is an epistemic phenomenon; but these philosophers thereby deny that *vagueness* involves indeterminacy. Indeterminacy is not our mere lack of knowledge as to a statement’s truth or otherwise. Rather, it is that there is no fact of the matter about which we *could* have knowledge. Certainly, it is impossible that we should ever come to know of an indeterminate statement that it is true, or that it is false; but moreover, this impossibility is not the mark of any cognitive defect on our part; it would not be overcome by any augmentation of our epistemic equipment. If ‘*S*’ is indeterminate, the facts fail to determine either that *S*, or that not-*S*. So no matter how much we might discover about the facts, we could never discover either that *S*, or that not-*S*.

Consider the following statement.

- (1C) There is carbon-based, multicellular life in the solar system other than on earth and its satellites.

This is a statement about the truth of which many are (justifiably) uncertain. However, this uncertainty does not arise from indeterminacy, but is due merely to a

lack of knowledge. Coming to know enough of the facts would resolve this uncertainty.

Now suppose I have a locker, and that I open it and find in it a strange object that looks like a misshapen spoon. Someone might say to me:

(1D) There is an etrier in your locker.

I might be uncertain as to whether this statement is true or false; but my uncertainty here would not arise from indeterminacy. My uncertainty would be the result of my unfamiliarity with the meaning of the word 'etrier'. Were I to discover that an etrier is a sort of rope-ladder, I would become certain that the statement is false.

Now let us try to find an example of a statement that *is* indeterminate. A much discussed type of apparently indeterminate statement arises on account of *vagueness*. I will not give a precise definition of 'vague'. Instead, following Williamson (1994: 2), I suggest that the best way of understanding what vagueness is, without prejudging the question of what the correct theoretical treatment of vagueness is, is by reference to some paradigm examples of vague expressions. Many kinds of expressions can be vague; but I shall be concerned only with *vague predicates*, and *vague singular terms*.³

Some examples of vague predicates are: 'is tall', 'is old', 'is short', 'is big', 'is bald' and 'is a heap'. One *mark* of the vagueness of these predicates is that they could all have *borderline cases*.⁴ To grasp the concept of a borderline case, consider, say, a series of one million dogs, beginning on the left with an (adult) Great Dane, and ending, on the right, with a Chihuahua, such that each dog is very slightly smaller than its predecessor. The first few dogs in this series are clearly large dogs. The last

³ As I will explain in Section 1.7, some believe that *properties*, as opposed to predicates, can be vague. One might believe that a predicate denoting a vague property can thereby itself truly be called 'vague'. Alternatively, one might claim that no predicate is vague, but that some predicates ('is bald' *etc.*) precisely denote vague properties. Someone holding this second view will I think be able to agree with most of what I say in the present chapter if s/he reads 'the property denoted by ...', where I have 'the predicate ...'. The same goes for persons believing that apparently vague singular terms ('Mt Everest' *etc.*) really are precise, but denote vague objects: s/he should read 'the object denoted by ...', where I have 'the singular term...'. However, I will in subsequent chapters explicitly assume a particular semantic understanding of the indeterminacy of indeterminate identity statements. (See Subsection 1.7.4.) That understanding is arguably an orthodox one; I do not offer any novel argument for it; but I propose to simplify my investigations by holding at least some things constant, and this is one of those things.

few are clearly not large dogs. However, there would appear to be no sharp dividing line between the large dogs and the dogs that are not large. There would appear to be at least some dogs in this series such that it is not clear that they are large dogs, and still not clear that they are *not* large dogs. These dogs are *borderline cases* of the vague predicate ‘is a large dog’.

We may use the term ‘borderline predication’ to refer to any predication of a vague predicate of one of its borderline cases. Now, borderline predications of vague predicates would appear to be examples of statements that are indeterminate. Where d is one of the borderline cases of ‘is a large dog’, consider the following statement.

(1E) d is a large dog.

This statement appears to be indeterminate, because the facts – facts about d ’s height, volume *etc.* – appear to fail to determine that d is a large dog, and to fail to determine that d is not a large dog. It would appear that, even if we knew all the relevant facts, about d ’s exact height, mass *etc.*, we should still be unable to say confidently either that d is a large dog, or that d is not a large dog – even though we take ourselves to be familiar with the meaning of ‘is a large dog’ – simply because ‘large’ is not a sufficiently precise term.

Let us now consider vague singular terms. One vague singular term is ‘the Southern Ocean’. That ‘the Southern Ocean’ is vague is apparent when we notice that some complex predicates of which it is the principal constituent are vague; *e.g.* ‘is in the Southern Ocean’. Since the Southern Ocean does not have sharp boundaries, some locations are borderline cases of this predicate. For such a location l , ‘ l is in the Southern Ocean’ seems to be indeterminate. Similar remarks apply to ‘Europe’, and for that matter to ‘Socrates’ arm’, since there are some portions of matter near Socrates’ shoulder which are borderline cases of ‘is a part of Socrates’ arm’.

There may be sources of indeterminacy other than vagueness. Under a certain interpretation of the theory known as ‘Intuitionism’, mathematical statements that have been neither proved nor disproved are indeterminate (in the sense in which I use ‘indeterminate’). Thus some would argue that Goldbach’s conjecture is indeterminate

⁴ Another mark of vague predicates is that they lend themselves to the construction of sorites paradoxes. I will discuss sorites paradoxes in Chapter 2.

– this being the thesis that every even number greater than 2 can be expressed as the sum of two prime numbers. (Flew 1984: 178) So, it may be that vagueness is not the only source of indeterminacy. But even if vagueness *is* the only source of indeterminacy, it is still not the case that ‘vagueness’ is synonymous with ‘indeterminacy’.

In Chapter 2 I shall scrutinise the concept of *indeterminacy* in more detail. But the above should be enough of an explanation of this concept for the purposes of the present chapter.

An important proviso is now in order. 1E would be an example of an indeterminate statement *unless* certain philosophers, *viz.* those who take the *epistemic view* of vagueness, are right. Proponents of this view believe that if we are unclear about whether, say, *d* is a large dog, this unclarity is due simply to our ignorance of the facts, be they facts about dogs, or facts about the usages or meanings of words. On the epistemic view, the facts determine of every dog either that it is a large dog, or that it is not a large dog. Defenders of this view deny that vagueness gives rise to indeterminacy. They would hold that there is a unique, sharp boundary in our series of dogs, which divides the (determinately) non-large dogs from the (determinately) large dogs. They would hold that there is a certain dog in this series which is the last of the determinately large dogs, the successor of which is the first of the determinately non-large dogs – despite there being but an imperceptible difference in size between these two dogs. This seems a strange view to take. It seems odd to assert that there really is this sharp boundary in the series. But defenders of the epistemic view have forceful arguments for just this assertion. I shall pursue the epistemic view in Chapter 2. However, I wish to *assume for the remainder of the present chapter* that this view is false, and that there are at least some series, pertaining to at least some vague expressions (for example the series of dogs just considered, pertaining to the vague expression ‘is a large dog’) which do give rise to indeterminate statements. I will put this assumption to the test, but not until Chapter 2.

1.2.5 ‘Indeterminate statement of diachronic personal identity’

Having defined ‘indeterminate statement’ and ‘statement of diachronic personal identity’ (‘SDPI’), we can see what it would be for SDPI «*P, Q, t, t**» to be indeterminate. The relevant facts, *e.g.* about the meanings of the words in the statement, and about the physical, psychological and other properties of *P* and *Q*,

would have to fail to determine either that P is numerically identical to Q or that P is not numerically identical to Q . Thus, even if we knew all the relevant facts, we should still be unable to say with confidence either that P is Q , or that P is not Q .

(Note: The above paragraph is an example of a passage that is implicitly schematic, the schematic letters in this case – ‘ P ’, ‘ Q ’, ‘ t ’ and ‘ t^* ’ – being singular terms. Moreover, some instances of the above passage-schema do not even make sense, since the substituent for ‘ $\langle P, Q, t, t^* \rangle$ ’ in such an instance will not even *be* an SDPI. In such an implicitly schematic passage as the previous paragraph, the instances that do not make sense can simply be disregarded: assume I intend virtually to utter, by such a schema, *only* those of the instances that it would make sense for me to utter.)

1.2.6 Possibility

I defined Personal Indeterminism as the claim that ‘there can be’ indeterminate SDPIs. I mean this to be read in the following way: ‘Neither logic, nor the laws of nature, rules out the possibility of some situation giving rise to indeterminate SDPIs.’ Thus, Personal Indeterminism will *not* be ruled out *merely* by the limited technical capabilities of humankind, but is ruled out, if at all, only by logic or the laws of nature.

1.2.7 Other locutions

(The following paragraph is schematic, ‘ S ’ being replaceable by indexical-free statements of the default context.)

To say that it is indeterminate whether S , is just to say that the statement ‘ S ’ is indeterminate. To say ‘it is determinate that S ’ or ‘it is determinately the case that S ’ or ‘determinately S ’, is just to say that the facts determine that S , which is just to say both (a) that S , and (b) that ‘ S ’ is not indeterminate.

(Locutions like the above, but for statements containing indexicals, or uttered in contexts other than the default context, would, I assume, admit of appropriate variations of the above explanation.)

1.2.8 A distinction

It is worth pointing out that Personal Indeterminism is not the same thesis as *reductionism* about persons, except on a fairly unconventional reading of

'reductionism'. Parfit (1984) *sometimes* uses the term 'the Reductionist View' to refer to Personal Indeterminism, in particular during the course of his Combined Spectrum Argument (237-241), which is certainly an argument for Personal Indeterminism, not for reductionism as conventionally so called. But I will never use 'reductionism' to denote Personal Indeterminism.⁵

1.3 INDETERMINATE STATEMENTS OF NON-PERSONAL DIACHRONIC IDENTITY

1.3.1

In his 1998, Simon Beck (§6) asserts: 'Actual phenomena like the gradual onset of conditions like senility or Huntington's corea already force us to accept that there can be cases in which it is indeterminate whether or not the concepts of person or same person apply'. But it is not clear that these cases 'force us' to accept this at all. I think many would believe that it is determinate that one and the same person continues to exist for the entire life of the human organism that, at the later stages of its existence, exhibits senility or *etc.*.

In fact, I do not think there are any real-life examples of obviously indeterminate statements of diachronic personal identity. There is not a convincing case for Personal Indeterminism without resort to hypothetical cases. (I will discuss some such cases shortly.)

1.3.2 Santa Trinita

However, there do seem to be real-life examples of statements of *non*-personal diachronic identity which it would be fairly uncontroversial to call indeterminate (given that we are assuming, for now, that there are *some* indeterminate statements). One such example is offered by Shoemaker.

In 1944 the Germans destroyed the four-century-old bridge of Santa Trinita in Florence. Six years later it was decided that it (?) should be rebuilt. On the original site there now stands a bridge of a design exactly like that of the original, constructed

⁵ See Johnston (1997: 261-262) and Garrett (1998: 38-39) for an outline of the various Parfitian usages
(footnote continued next page)

by Renaissance techniques, and built in part with the original stones (each standing in its original place), in part with new stones taken from the original quarry. The facts are all clear, but how are we to answer the question “Is the present bridge at Santa Trinita the very bridge that spanned the Arno four hundred years ago?” One can imagine one person saying “This is a modern copy of a Renaissance bridge that once stood here,” and another, *equally cognizant of the facts*, saying “This bridge has been the pride of Florence for four centuries.” [My emphasis] (Shoemaker 1963: 28-29)

This appears to be a real-life example of an indeterminate statement of non-personal diachronic identity: ‘The present bridge at Santa Trinita is the very bridge that spanned the Arno four hundred years ago.’ This statement appears to be indeterminate, in that the facts fail to determine either that we are dealing with the same bridge, or that we are dealing with a new bridge.⁶ One can imagine a dispute over whether this statement is true or false, between two people who are nevertheless fully agreed on all the facts of the bridge’s(s’) history, construction and composition. Such a dispute, as Shoemaker points out, could reasonably be characterised as just being ‘about words’. (1963: 29)

1.3.3 The Inanimate Spectrum

Suppose one does not agree that the above example gives rise to an indeterminate diachronic identity statement. Suppose, *e.g.*, that one thinks the bridge at Santa Trinita now is clearly *not* identical to the bridge there four hundred years ago. One might yet be convinced, however, *via* a more thorough method, that there can be statements of non-personal diachronic identity that are indeterminate. In part to set the scene for Section 1.4, I will now describe a *spectrum* of hypothetical cases, which I will call ‘the Inanimate Spectrum’. That these are not actual cases will not handicap my argument, as I am merely arguing that there *can* be statements of non-personal diachronic identity, in the sense that some possible situation – one which violates neither logic nor the laws of nature – would render such a statement indeterminate; and the cases I will consider violate neither logic nor the laws of nature.

of the phrase ‘Reductionist View’.

⁶ Another good example of an apparently indeterminate non-personal diachronic identity statement can be found in Parfit’s example of a club which dissolves but is formed again at a later date. (Parfit 1984: 213)

Suppose, then, that we have a particular building, b , which is made entirely of brick, existing at a time t . Building b is 3 metres high, 3 metres wide and 3 metres deep. Consider now a spectrum of hypothetical cases, $C(1)$, $C(2)$, ... , $C(1,000,000,001)$, which are to be envisaged not as occurring one after the other but rather as being alternative scenarios, any one of which *might* happen. In each of these cases a particular event happens to b . In $C(1)$, nothing happens at all. In $C(2)$, one billionth of the material composing b is removed, ground to dust and scattered in the ocean; then the space that was taken up by the removed part is filled with an entirely different, modern synthetic material. In general, in case $C(n)$ (where $1 < n \leq 1,000,000,001$), all of the matter composing b that would have been removed in $C(n-1)$ is removed, and in addition one extra billionth of b 's matter is removed (all removed matter being removed at once, rather than in stages); all the removed material is ground to dust *etc.*; and all spaces that were filled by the removed material are then refilled with the synthetic material. In $C(1,000,000,001)$, all of the original building is removed, ground to dust, and scattered in the ocean, and is then replaced with a building made entirely of the synthetic material. In the following discussion ' $b(n)$ ' refers to the building in place at the conclusion of case $C(n)$.

1.3.4 The Inanimate Spectrum Argument

Determinately, $b(1)$ is one and the same building as b . Nothing has been done to b at all in case $C(1)$. It seems, too, that $b(2)$ is determinately one and the same building as b , since the removal and subsequent replacement of only one billionth of b 's material with synthetic material would not seem a significant enough change to cause the building to go out of existence. At the other end of the spectrum, it is surely determinate that $b(1,000,000,001)$ is *not* one and the same building as b . In this extreme case, we originally had a brick building; then we had a building made of totally different material erected on the spot where the old building used to stand.

Suppose that statements of non-personal diachronic identity cannot be indeterminate. In that case, for every n ($1 \leq n \leq 1,000,000,001$), either it is determinate that b is $b(n)$, or it is determinate that b is not $b(n)$. Given that it is determinate that b is $b(1)$, and it is determinate that b is not $b(1,000,000,001)$, there must be some n ($1 < n \leq 1,000,000,000$) such that in $C(n)$ it is determinate that b is $b(n)$, and in $C(n+1)$ it is determinate that b is not $b(n+1)$. But for any n , the difference between $b(n+1)$ and $b(n)$ is virtually imperceptible.

It is very difficult to believe that the facts determine of a particular case in the Inanimate Spectrum that it is the last of the *b*-preserving cases, and its successor the first of the *b*-destroying cases. Our language – the word ‘building’, ‘the same as’, *etc.* – just does not seem precise enough for there to be any such crucial point in the series. It seems highly implausible to suppose that there is this sharp boundary. This seems particularly difficult to believe when we ask ourselves: What could we possibly do to discover its exact location in the Spectrum? Suppose we know everything about the physical composition of the buildings, their size, shape *etc.*; we know their history; we can physically inspect them in as much detail as we like; and we are fully competent users of our language. Yet it still seems that we would not know exactly at which point this boundary allegedly determinately lies; and there would seem to be nothing else we could do to try to figure out where the alleged sharp boundary is. Trying to figure this out would seem pointless – a futile, vacuous exercise. Our inability to locate a sharp boundary does not logically entail that there is no sharp boundary. However, the only plausible explanation for this inability in *this* case seems to be that our concept of “building”, or of “the same building”, is not so precisely refined as to yield a definite answer, for every possible pair of buildings, to the question of whether or not they are the same building.

It seems there must be some cases in the Spectrum such that the facts do not determine either that they are *b*-preserving cases, or that they are *b*-destroying cases. For any such case $C(k)$, the facts determine neither that $b(k)$ is *b*, nor that $b(k)$ is not *b*. The statement (of the form) ‘ $b(k)$ is *b*’ is thus indeterminate; and it is a statement of non-personal diachronic identity (‘ $b(k)$ ’ and ‘*b*’ having been defined by reference to the times “at the conclusion of case $C(k)$ ”, and *t*, respectively). Because the cases in the Inanimate Spectrum are *possible* (violating neither logic nor the laws of nature), we can conclude that there *may* be indeterminate statements of non-personal diachronic identity. Let us call this argument ‘the Inanimate Spectrum Argument’.

There are certain philosophers who would object to the Inanimate Spectrum Argument, *viz.* defenders of the epistemic view of vagueness, which I mentioned earlier. Again, let us put their challenge to one side for now; I shall deal with it properly in Chapter 2.

1.3.5 Evans's argument does not target *statement indeterminacy*

Some who do not take the epistemic view may still want to reject the conclusion of the Inanimate Spectrum Argument, because they have been persuaded, by Gareth Evans's (1978) much-discussed paper, that there can be no indeterminate *identity* statements.⁷ However, I would refer anyone taking this view to David Lewis (1988), who argues convincingly that Evans did not intend to argue against the possibility of indeterminate identity statements, and that Evans's argument, on *that* (mis)interpretation, would in any case be fallacious. I will not repeat Lewis's points here. I will assume from now on that the only basis upon which one might want to object to the Inanimate Spectrum Argument is the epistemic view of vagueness.

I shall return briefly to Evans's argument in Subsection 1.7.2.

1.4 THE COMBINED SPECTRUM ARGUMENT

1.4.1

I will now examine Derek Parfit's argument for the possibility of indeterminate statements of diachronic *personal* identity (indeterminate SDPIs). We shall see that this argument is based on a spectrum of hypothetical cases very similar to the Inanimate Spectrum.

Parfit's argument for Personal Indeterminism – the thesis that SDPIs may be indeterminate – is effected by a thought experiment, or rather a spectrum of thought experiments, that he calls 'the Combined Spectrum'. (Parfit 1984: 236) Parfit sets the scene for the Combined Spectrum by first asking us to entertain two other 'spectra' of imaginary cases: the 'Physical Spectrum', and the 'Psychological Spectrum'. (1984: 231-6) However, we need only to discuss the Combined Spectrum here, as all the elements of Parfit's argument for Personal Indeterminism are contained within his discussion of that Spectrum alone. (We will touch on the Physical and Psychological Spectra in Section 1.5.)

⁷ See Salmon (1981: 244ff) for a similar argument, independently proposed.

1.4.2 Differences with Parfit's presentation

My presentation of the Combined Spectrum, and of the argument which Parfit bases on it, will differ from Parfit's presentation in three ways.

Firstly, where Parfit speaks of "me" and "Greta Garbo", I will speak of "*P*" and "*Q*".

Secondly, while Parfit discusses the Combined Spectrum *after* he has already argued against the *Simple View* of personal identity (the view "that DPI facts do not supervene on non-DPI facts") (see Chapter 3), I present the Combined Spectrum Argument first, and will introduce the Simple View afterwards, as a possible source of resistance to the Combined Spectrum Argument. In Chapter 3, I explain the Simple View, and its relation to Personal Indeterminism, and assess whether the Simple View is plausible. But note that the Simple View is an ontological thesis and is *not* the same thing as the denial of Personal Indeterminism, even though the two are closely related, in a way that will become clear in Chapter 3.

Thirdly, while Parfit, in a certain sense, describes the *operations* in the Combined Spectrum from a physical standpoint (even though it is not the same as his Physical Spectrum), I shall describe the operations in the Combined Spectrum from a jointly physical and psychological standpoint.

1.4.3 The Combined Spectrum

The Combined Spectrum, then, can be described as follows. Suppose there is a person *P* (existing at a time *s*) and a person *Q* (who died earlier than *s*), who are physically and psychologically⁸ very different from one another. Suppose that scientists recorded all of *Q*'s physical and psychological characteristics at time *r* prior to *Q*'s death, in perfect detail. We suppose that they have the ability to create, out of new matter, cells which have or sustain exactly all of the properties had or sustained by *Q*'s cells at *r*, *i.e.* to create a perfect replica of all or part of *Q*. Imagine a spectrum of hypothetical operations any one of which might be performed on a person *P* at a certain time *t* later than *s*. The operations are not to be imagined as a series of operations occurring one after another, but are rather to be envisaged as a spectrum of alternative hypothetical operations any one of which might be performed on *P* at *t*. In each operation, a certain portion of *P*'s cells are destroyed and replaced with replicas of *Q*'s cells, and a certain

⁸ I take it that *behavioural* tendencies count as psychological features.

number of P 's psychological characteristics (*e.g.* memories) are changed, so that P becomes more like Q in these respects. (That psychological change will occur in virtue of the cell replacement, if psychological features depend on physical ones; but suppose, if they do not so depend, that they can be altered in some other way.) At one end of the spectrum (the "leftmost" extreme), P is placed on the operating table, but nothing is done to P at all. Here there is psychological and physical continuity between P as s/he is immediately before the operation, and the post-operation person as s/he is immediately after the operation. At the other end of the spectrum (the "rightmost" extreme), P is physically destroyed, and an entire physical and psychological replica of Q is placed on the operating table instead. Here there is neither psychological nor physical continuity between P as s/he is immediately before the operation, and the post-operation person as s/he is immediately after the operation. Each operation, other than the "leftmost" operation, includes all the physical and psychological changes that happen in the operation immediately to its "left", but also includes *very slightly* more physical and more psychological changes than the operation immediately to its "left". We suppose that surgeons can do all of these operations while keeping alive whatever *body parts* are on the operating table. There are no surgical "complications" (merely philosophical ones). A living person is on the operating table before the operation; and a living person is on the table after the operation. (Parfit 1984: 236-7)

1.4.4 The Combined Spectrum Argument: the basic argument

To begin with, observe that at one extreme of the Spectrum, where nothing is done to P at all, it is determinately the case that P is identical to the person on the table after the operation. In the operation at the other end of the Spectrum, it is determinate that P is *not* identical to the person on the table after the operation, since here P is physically annihilated, and replaced by a replica of Q .

Suppose that Personal Indeterminism is false. Then if ' S ' is an SDPI, the facts determine either that S , or that not- S . It follows that there is an operation O in the Combined Spectrum such that the facts determine that the person on the table after operation O is P , but such that, in the operation O^* that is one to the "right" of O , in which the physical and psychological changes are only very slightly greater than in O , the facts determine that the person on the table after O^* is *not* P . In Parfit's words, 'There must be some critical set of the cells replaced and some critical degree of

psychological change, which would make all the difference.’ (1984: 238-9) This critical percentage would be just sufficient to result in the termination of *P*. (If *P* is not identical to the person on the table after the operation, then *P* does not exist after the operation, since there is presumably no-one else existing after the operation to whom *P* might be identical.)

The structure of the Combined Spectrum is clearly very similar to that of the Inanimate Spectrum. In both, the denial of indeterminacy entails the existence of a sharp boundary between the cases that would make the relevant identity statement true, and the cases that would make it false.

Parfit makes two points about why it is difficult to believe that there is a sharp boundary in the Combined Spectrum. His *first point* is that it is difficult to believe ‘that the difference between life and death could consist in any of the very small differences’ between neighbouring cases in the Spectrum. (1984: 239) He argues:

We are inclined to believe that there is *always* a difference between some future person’s being me, and his being someone else. And we are inclined to believe that this is a *deep* difference. But between neighbouring cases in this Spectrum the differences are trivial. It is therefore hard to believe that, in one of these cases, the resulting person would quite straightforwardly be me, and that, in the next case, he would quite straightforwardly be someone else. [Parfit’s emphasis] (Parfit 1984: 239)

Parfit’s *second point* about what makes it difficult to believe that there is a sharp borderline is that ‘we could never have any evidence where the borderline would be’. (1984: 239)

On these grounds, Parfit argues that the claim that there is a sharp boundary in the spectrum, between the *P*-preserving operations and the *P*-destroying ones, is an extremely implausible one. The denier of Personal Indeterminism is committed to this claim. Parfit thinks that this claim is more implausible than Personal Indeterminism. Therefore we should conclude, argues Parfit, that Personal Indeterminism is true: we should reject the notion ‘that our identity must be determinate’. (Parfit 1984: 238)

What should we make of Parfit’s argument?

1.4.5 Comparison with the Inanimate Spectrum Argument

Firstly, observe that, given their structural similarity, the only relevant differences between the Inanimate Spectrum and the Combined Spectrum arise from the fact that the latter concerns persons, while the former does not. One relevant difference just is the fact that the latter concerns persons and the former does not. Another relevant difference is that, while in the Inanimate Spectrum the differences between cases were described in wholly *physical* terms, by referring to the quantity and physical qualities of building materials, in the Combined Spectrum the differences between cases were described in both *physical and psychological* terms. This is because the respects of difference that are *relevant* to the identity of buildings are unarguably all physical respects; but the respects of difference that are relevant to personal identity arguably include both psychological respects and physical respects. Of course, one might argue that psychological differences really just reduce to, or depend on, physical differences; but one can frame the Combined Spectrum Argument in a way that does not require one to presuppose that all the potentially relevant respects of difference reduce to or depend on physical respects, if one *specifies* in one's description of the spectrum that the difference between the products of any two neighbouring operations is slight in all physical *and* psychological respects.

Secondly, like the Inanimate Spectrum Argument, the Combined Spectrum Argument derives its force from the supposed implausibility of postulating a sharp boundary. In the Inanimate Spectrum, one of the things that makes it especially difficult to believe in a sharp boundary is that there seems to be no way of finding out where it is. Parfit raises an analogous point, his *second point*, in the Inanimate Spectrum Argument. But he also raises another point, his *first point*, which appeals to the apparently 'deep' difference between 'being me' and 'being someone else'. Parfit's *first point* is specific to *personal* identity: it does not have an analogue in the Inanimate Spectrum. We are not inclined to believe there is anything especially deep about the identity of buildings.

Presently I will discuss Parfit's *second point*. Then I will discuss his *first point*.

1.4.6 Parfit's *second point*

In the Inanimate Spectrum, it was difficult to see how, as observers of the building(s), we might go about figuring out where an alleged sharp boundary would lie. And

considered from the perspective of an observer, *i.e.* from a third-person perspective, the Combined Spectrum seems similar in this regard. It would appear futile, from that perspective, to try to figure out where an alleged sharp boundary would lie. From a third-person perspective, it seems that the difficulty of locating a sharp boundary in the Combined Spectrum (or obtaining evidence about the location of such a boundary, as Parfit puts it) is just as great as in the Inanimate Spectrum.

It might be thought that we could simply ask the patient emerging from any of the hypothetical operations, who s/he is. But in any pair of neighbouring cases, the psychological qualities (memories, intentions *etc.*) of the post-operation patient in one case, and the psychological qualities of the post-operation patient in the other case, differ only very slightly. Any reliable judgment formed, after the operation, by the post-operation patient him/herself about his/her own personal identity would have to be formed on the basis of *some* information or other – where ‘information’ is meant here in a broad sense, so as to encompass, besides observations of the outside world and of the patient’s own body, the having by the patient of memories and/or other psychological qualities, experiences, *etc.*. The patient him/herself would appear to be in no better position to judge his/her own identity than we are as outside observers. (In Chapter 3 I will address an argument that persons can have *direct* knowledge of their diachronic identity, rather than *via* memories *etc.*. But let us ignore this possibility for the time being.) It is difficult to believe that, after one operation, *O*, the post-operation person will answer with conviction ‘I am *P*’, but after another operation, *O**, such that the memories *etc.* of the post-operation person differ only very slightly from those of the post-operation person after *O*, the post-operation person will answer with conviction ‘I am *Q*’ or ‘I have only just come into existence’. We would expect, rather, a zone of central cases in the Spectrum in which the post-operation person is unclear about his/her own identity. This zone, in which identity is unclear from a first-person perspective, would be based on the gradualness of psychological difference across the Spectrum – just as the zone of unclarity about the patient’s identity from a third-person observer’s perspective is based on the gradualness of physical (and behaviourally manifested psychological) difference across the Spectrum.

It appears, then, that one would have just as much difficulty locating an alleged sharp boundary in the Combined Spectrum as in the Inanimate Spectrum, whether one is an observer, or the post-operation patient him/herself. In both Spectra, this difficulty would persist despite one’s knowing all the relevant physical – and, in

the case of the Combined Spectrum, psychological – facts, and despite one’s being a fully competent user of one’s language.

So much for Parfit’s *second point*. Let us now consider his *first point*.

1.4.7 Parfit’s *first point*

Parfit’s *first point* appeals to the difficulty of believing that such a minuscule difference as that between two neighbouring cases in the Combined Spectrum could make the difference between life and death; or, put differently, that it could make it true that in one case, a future person is ‘quite straightforwardly’ me, and in another case, but imperceptibly different from the first, a future person is ‘quite straightforwardly’ someone else. (Parfit 1984: 239) Parfit states that we are inclined to believe that the difference between a person’s being me, and his/her being someone else, is a deep difference. But the differences between any two neighbouring cases in the Spectrum are trivial.

It was noted that Parfit’s *first point* is specific to *personal* identity. Now, it may seem odd that Parfit appeals to the impression that diachronic personal identity (DPI) is ‘deep’, since elsewhere he claims that DPI ‘just involves’ certain more particular facts, *e.g.* facts about what memories a person has, and that DPI ‘is not what matters’. (1984: 216-217) It seems to be of the very essence of his so-called “Reductionist View” of personal identity to deny that DPI is ‘deep’. However, Parfit in his *first point* does not *claim* that DPI is ‘deep’. Rather, he claims that we are inclined (rightly or wrongly) to think that DPI is deep. If we are inclined to this view, argues Parfit, then we should find it difficult to believe that the ‘deep’ difference between DPI holding, and DPI not holding, could hinge on such an apparently trivial difference as that between two neighbouring cases in the Combined Spectrum. But if we are *not* inclined to this view – *i.e.* if we think that DPI facts are no deeper than facts about, say, the identity of buildings – then there is in any case no motivation for us to resist the conclusion that DPI statements can be indeterminate, just as the identity of buildings can be indeterminate. Of course, this assumes that one accepts the Inanimate Spectrum Argument. But I will address those who would reject that argument in Chapter 2.

1.4.8 Two minds?

There is one objection to the Combined Spectrum Argument which I will deal with immediately.

Madell questions the notion that ‘chunks from different minds could be somehow welded together’. And he argues that even if they could be welded together, ‘it is quite unclear that the outcome would be as Parfit supposes.’ (1985: 107) Rather than ending up with a spectrum of post-operation “hybrid” persons each of whom is to a greater or lesser degree like *P* or like *Q*, Madell suggests that ‘[i]t seems more natural to suppose that were it possible to combine some of your mind with some of mine, what we would have is something like two minds in the one skin, a pretty startling case of multiple personality perhaps.’ (1985: 107)

This suggests two objections. One is a charge of impossibility: that the operations could not be carried out, since different minds could not be ‘welded together’. I shall address that charge in Subsection 1.4.9. The other charge is that even if such welding could be carried out, the result would more likely be a sharing by multiple minds of a single skull, rather than the hybrid single mind suggested by Parfit’s discussion of the spectrum.

Though it is questionable that ‘two minds’ is what we would end up with if such operations were carried out, the Combined Spectrum Argument would work even we *did* end up, in some cases, with two minds. Consider that the “leftmost” operation leaves *P* completely intact. Assuming he was one-minded to begin with, he will surely remain so after this null operation. At the rightmost extreme of the Spectrum, *P* is destroyed utterly and replaced with a perfect replica of *Q*. Assuming, then, that *Q* was one-minded, his replica surely will be too. So if Madell thinks that, assuming the operations to be possible, some of them would result in two-minded persons, or two person-minds in one body, there would remain the difficulty for him of justifying the postulation of a sharp boundary between the cases in which there is one mind, and the cases, in the middle, in which there are two. It is just as implausible to posit *such* a sharp boundary as it is to posit a sharp boundary in the one-minded, hybridising conception of the Spectrum. Thus, if Madell is right about what would happen if we carried out the operations, it would seem to be indeterminate, for some operations, whether there is one mind or two; for a tiny psychological, behavioural and physical difference would surely not make the difference between whether there is *determinately* one mind or *determinately* two.

Madell might rejoin that it can never be indeterminate whether there can be one mind or two. But there is no reason to think this is so. There have surely been real life cases in which it is plausible to construe a particular brain as supporting an indeterminate number of minds – perhaps one, perhaps two. Moreover, I suggest that, if Madell merely means, by ‘two minds’, just ‘two personalities’ then it is especially easy to believe it might be indeterminate how many “minds” there are. But if he means something like ‘two spheres of conscious awareness’, then I suggest it is much less plausible to think any of the cases must generate two “minds”.

Motivation for thinking the number of minds to be a necessarily determinate affair might perhaps come from Madell’s “Simple View” of personal identity: I shall address the Simple View in Chapter 3.

So, the most plausible picture of the Combined Spectrum, *if* we are to think of some operations in it as generating two minds sharing one skull, is one in which there is a fuzzy boundary between the cases in which there is one mind and the cases in which there are two (actually two fuzzy boundaries, one at each end of the two-minded zone). So there would be cases for which it is indeterminate whether the post-operation skull contains one mind or two. (There might be a better word than ‘contains’ here, but this does not affect the argument.)

Let us grant that all the one-minded post-operation skulls would belong to persons who are determinately identical or determinately non-identical with *P*.

Suppose we think each two-minded case (*i.e.* case in which the post-operation skull “contains” two minds) must involve two post-operation persons, one per mind. Consider cases where it is indeterminate whether there is one mind or two in the post-operation skull. Then for a mind *a* and a mind *b*, it is indeterminate whether $a=b$. Call the person who after the operation has mind *a* ‘*A*’, and the person who after the operation has mind *b* ‘*B*’. Then, it will be indeterminate whether $A=B$.⁹ This would be a case of (an) indeterminate synchronic personal identity (statement). But we could easily establish there to be (an) indeterminate diachronic identity (statement) between at least one of *A* or *B*, and the pre-operation person *P*. For suppose *P* is determinately *A*. Then since it is indeterminate whether $A=B$, it is indeterminate whether $P=B$.

⁹ Cf. Shoemaker’s (1984: 145-146 n. 5) example of Alpha and Beta Hall, two structures joined by a narrow walkway. It is indeterminate whether they together constitute one building, or two. So pointing to Alpha Hall, we say, ‘call this building *a*’, and pointing to Beta Hall, we say ‘call this building *b*’. Then it is indeterminate whether $a=b$.

Suppose P is determinately B . Then, by similar reasoning, it is indeterminate whether $P=A$. So, assuming there are no indeterminate SDPIs, P is neither A nor B . Then either (a) the boundary between the left lot of one-minded cases, and the two minded cases, is the boundary between the P -preserving and the P -terminating cases, or (b) the boundary between the one- and two-minded cases is not significant in this respect. If (a), since there is no sharp boundary between the one- and two-minded cases, there is no sharp boundary between the P -preserving and the P -destroying cases, thus generating indeterminate SDPIs. But if (b), the postulation of two-minded cases has not helped the opponent of Personal Indeterminism to avoid commitment to an implausible sharp boundary with respect to DPI.

Suppose on the other hand that each post-operation body that contains two minds is construed as belonging to one person. The middle cases would be the ones resulting in two minds. Then it would be implausible to draw a sharp boundary between the cases in which P is the two-minded post-operation person, and cases in which P is not that person.

1.4.9 Alleging impossibility

Madell (1985), besides questioning whether two minds might be combined into one, apparently also questions whether two minds might be combined in *any* sense – whether ‘chunks from different minds [might] be somehow welded together.’ (107)

Madell (206) seems to recognise that while the operations in the Combined Spectrum might forever remain technically impossible, this kind of impossibility alone would not undermine the Combined Spectrum Argument: Personal Indeterminism alleges just that indeterminate SDPIs are ruled out neither by logic nor by the laws of nature. But he notes with interest Parfit’s (1984: 255) admission that ‘it seems likely that it would never be possible to divide the lower brain, in a way that did not impair its functioning.’ That is, although the two hemispheres of the upper brain might be divided, Parfit admits that we might never be able to divide the lower brain – but he thinks ‘[t]his impossibility is merely technical’. But Madell wonders on what grounds we are justified in thinking this apparent technical impossibility is *merely* technical, and not indicative of a deeper conflict with the laws of nature.

Parfit’s admission that we might never be able to divide the lower brain without ‘impair[ing] its functioning’ is made during a discussion of personal “*division*”, not of the Combined Spectrum. But it does seem that, *if* we regard the

question of what proportion of the lower brain of the post-operation person was also a part of the lower brain of the pre-operation person (*P*), and the question of what proportion of psychological features supported by this lower brain matter have changed, or stayed the same, in the course of the operation, as questions that are relevant to DPI, then we should have to suppose that varying proportions of the lower brain of *P*, and of those psychological features, could be replaced during the operation. If it were not possible to take out, say *half*, or a *quarter*, of the lower brain of *P*, and replace that half or quarter with new functioning brain-matter – if the lower brain could be replaced only as a whole – then the denier of Personal Indeterminism might seem justified in positing a sharp boundary between the neighbouring operations *O* and *O** such that the lower brain of *P* stays in place in *O* but is removed in *O**. And if *P*'s lower brain is not replaced in *any* operation, perhaps one might claim that *P* survives all the operations, and thus evade the challenge of drawing a boundary altogether. Thus, it seems, at least initially, that if function-preserving division of the lower brain, and of the psychological features it supports, is ruled out by the laws of nature, then the Combined Spectrum Argument fails.

Madell (1985: 106-107) points out that one argument seemingly offered by Parfit for the claim that the impossibility is *merely* technical is this: 'Since our psychological features depend on the states of our brains, these imagined cases are only technically impossible.' (Parfit 1984: 238) Madell thinks this an ineffective argument: 'even if ... every mental state and event is fully determined by states of and events in our brains ... this offers no support for Parfit's view of psychological states as in principle extractable and replaceable by others from other persons. Even the physical determinist might well regard the suggestion that, say, Parfit's interest in Buddhism could be taken out of Parfit's mind and replaced by Greta Garbo's Longing to be Alone as nothing but a fantasy.' (1985: 107)

Madell seems, however (though it is not completely clear) to be overstating what is required by the Combined Spectrum Argument. If psychological features depend on brain-states, it need not be possible to target *particular* memories, intentions *etc.* for replacement. It is just implausible to suppose that, in two operations which differ only in that the second replaces only one more neuron (or, if you like, only one more neuron-neuron connection), there could be *any sufficiently great* psychological difference between the post-operation persons such that one is determinately identical with *P* and the other is determinately non-identical with *P*.

It is reasonable to think that psychological features do depend on brain-states. But even if psychological features do not depend on brain-states, this would not by itself undermine the Combined Spectrum Argument. The mere non-dependence of something on the physical does not entail that it does not admit of degrees. Even if psychological qualities do not depend on physical ones, and even if it is technically impossible to alter psychological features unless one can do so *via* brain-tampering, it still is plausible to think psychological features can differ only by slight increments, across a spectrum, in a way that contravenes neither logic nor the laws of nature; for I can imagine easily enough, *e.g.*, the deletion of just a few memories, or apparent memories from my mind, and the addition of just a few new ones (or, alternatively, a slight tweaking of one or two of my existing memories), and I can do this without presupposing anything about whether my brain would have also to change physically at all. Of course, that I can *imagine* this may not be enough to establish possibility; but it seems even that in real life, people really *do* lose memories *gradually*, not just in cases of progressively worsening Alzheimer's disease, but to some extent in almost all people as they get older; and people really *do* sometimes gain false memories. It is implausible to think that the very laws of nature prohibit any very extensive memory replacement. And these claims, about the gradualness with which memory loss or replacement might occur, do not appear to presuppose any supervenience of memories on the physical. If we discovered that psychological properties supervened on the qualities of some immaterial substance, for example, that discovery should not in itself give us any special reason to think that certain degrees of psychological change are prevented by the laws of nature.

What about dividing the lower brain? Is it plausible to think that the *laws of nature* or *logic* prevent it, or whatever psychological features it supports, from being divided? It is not. We have no conclusive evidence for this claim, of course; to some extent, this *is* a matter of speculation, given our comparative ignorance about the brain. But it is unlikely that some deep facet of nature prevents this kind of division. It is rather more likely to be an accident of the particular design of our brain.

Further, it is implausible to suppose that I should continue to exist even were my entire upper (*i.e.* non-lower) brain replaced. Why should we think the lower brain to be the seat of personal identity? So, assuming even that the operations in the Spectrum do not tamper with *P*'s lower brain at all, it is still implausible to draw a sharp boundary between the cases in which enough of *P*'s upper brain remains intact

to preserve *P*, and the cases where so much of *P*'s upper brain has been replaced as to cause a new person to come into existence, but one who happens to inherit the lower brain that was *P*'s.

1.4.10 Two more ways of trying to resist the argument

Assuming that one has accepted the Inanimate Spectrum Argument, is there any way to resist the Combined Spectrum Argument, apart from the two ways just rejected? To do this, one would apparently have to claim that there is something appropriately special about *personal* diachronic identity. Here are two (not necessarily mutually exclusive) ways in which one might want to make such a claim.

- (Way 1) Because some form of the so-called Simple View about personal identity is true, the postulation of a sharp boundary cannot be shown to be incorrect *via* the Combined Spectrum Argument.
- (Way 2) The first-person perspective that is distinctive of personal identity renders incoherent the notion that DPI statements can be indeterminate.

Consider Way 1. Parfit himself thinks that someone who believes that a person is a 'separately existing entity' may resist the Combined Spectrum Argument. (1984: 237-238) The thesis that persons are 'separately existing entities' is a form of the Simple View of DPI. Parfit defends the Combined Spectrum Argument from Way 1 by arguing, elsewhere, that persons are not "separately existing entities". (1984: 223-6) I will not address the question of whether Way 1 is actually feasible until Chapter 3. In Section 1.6 (in the present chapter), I will address the question of whether Way 2 is a feasible way of blocking the Combined Spectrum Argument.

1.4.11

It will be helpful at this point to summarise the overall structure of the case for Personal Indeterminism presented in this chapter.

- (Premiss 1) The Inanimate Spectrum Argument shows that there can be indeterminate statements of non-personal diachronic identity.
- (Premiss 2) If the Inanimate Spectrum Argument shows that there can be indeterminate statements of non-personal diachronic identity, then the Combined Spectrum Argument shows that there can be indeterminate statements of diachronic personal identity.
- (Conclusion) There may be indeterminate statements of diachronic personal identity.

In Section 1.6, I will test the resistance of Premiss 2 to Way 2. In Chapter 3, I will test the resistance of Premiss 2 to Way 1. In Chapter 2, I will discuss arguments for the epistemic view, which doctrine would deny indeterminacy in both the Inanimate and the Combined Spectra. If the epistemic view is correct (and I will argue that it is not), then Premiss 1 would be undermined.

Presently, however, I will examine a critique of Parfit's reasoning by Goodenough (1996), which does not fit neatly into the above scheme.

1.5 GOODENOUGH'S CRITIQUE OF PARFIT

1.5.1

Goodenough claims that Parfit misuses sorites reasoning in his deployment of the so-called Physical, Psychological and Combined Spectra. (Parfit 1984: 231-43) Two claims are involved here. First, it is claimed that Parfit 'make[s] use of [s]orites-based reasoning'. (Goodenough 1996: 113) Second, it is claimed that making use of sorites-based reasoning, at least in the way that Parfit does, is a bad thing: it constitutes weak and invalid argument. Says Goodenough, 'the Sorites demon ... tends to undermine a great deal more than was intended ... [and an] example of this problem ... can be found in the work of Derek Parfit.' (Goodenough 1996: 113)

Goodenough also alleges that Parfit, in his rejection of sharp dividing lines in his Spectra, commits himself to an unacceptable conclusion.

I will argue, firstly, that Parfit does *not* use ‘Sorites-based reasoning’. Secondly, I will note that Parfit’s rejection of sharp dividing lines does not commit him to the unacceptable conclusion – though I will not defend this claim fully till Chapter 2.

1.5.2 The Physical Spectrum

So far we have discussed only the Combined Spectrum, since in terms of establishing Personal Indeterminism, the other two Spectra are superfluous. However, to understand Goodenough’s critique, we will need to familiarise ourselves with Parfit’s Physical Spectrum.

In the Combined Spectrum, the changes effected by the hypothetical operations are both physical and psychological. The Physical Spectrum is similar, but with change occurring only on the physical dimension. Again we imagine a spectrum of hypothetical operations any one of which might be performed on a patient *P* at a certain point in time. In each operation in the Physical Spectrum, a certain amount of the matter constituting the patient *P* is replaced with new, but *qualitatively identical*, matter.¹⁰ At the “near” end of the Spectrum, a very little of *P*’s matter is replaced. At the “far” end, all of *P*’s matter is destroyed and replaced with new matter. In all cases the person on the table after the operation is psychologically continuous with the person on the table before the operation. (Parfit 1984: 234-6)

Parfit frames his Physical Spectrum in response to an argument of Bernard Williams’s (1970) against the so-called Psychological Criterion of personal identity. We will meet this argument of Williams’s in Subsection 3.3.3, under the label ‘Argument Two’, and will consider critiques of it (though not, explicitly, Parfit’s¹¹) there. For now, we need just to note the following. The Psychological Criterion is a particular view about what are the necessary and sufficient conditions for diachronic personal identity to obtain. Within a certain framework of assumptions, it competes with the so-called Physical Criterion of personal identity. We are not interested here in this particular debate. Our interest is in whether or not there can be indeterminate SDPIs. However, Goodenough’s interpretation of Parfit’s use of the Physical

¹⁰ See Note 2.

¹¹ In brief: Williams (1970) (i) rejects Personal Indeterminism and (ii) endorses a physical, non-psychological criterion of personal identity. Parfit (1984: 234-236) argues *via* his Physical Spectrum (footnote continued next page)

Spectrum is relevant to the task at hand. As we shall see, if Goodenough is right, Parfit commits himself to a form of reasoning that leads to highly contentious conclusions; and if Goodenough is right, moreover, these conclusions might be escaped only by accepting what we have earlier denied in arguing for Personal Indeterminism: that there is a sharp dividing line in the Combined Spectrum. Let us see how this is so.

1.5.3 Goodenough's interpretation of Parfit

Goodenough interprets Parfit as using a sorites-style argument to argue against the Physical Criterion. Believers in the (pure) Physical Criterion hold that diachronic identity 'just involves the physically continuous existence of enough of a brain so that it remains the brain of a living person' (Parfit 1984: 209-210) – regardless of the extent and quality of the psychological connections between the person possessing such a brain-portion at one time, and the person possessing it at another time. Thus, the Physical Criterion would have it that a person will continue to exist (*i.e.* will still be identical with some existent person) even in case of complete psychological erasure (of all memories, intentions and other psychological "contents"), providing an appropriate portion of the person's brain continues to exist. Goodenough *thinks* that Parfit argues against the Physical Criterion as follows.

"The Replacement Argument"

Let N be the number of cells in my brain and body and let P be my personal identity.

- (R1) N sustains P.
- (R2) If N sustains P, then (N-with-1%-cells-substituted) sustains P.
- (R3) If (N-with-x%-cells-substituted sustains P) then (N-with-x+1%-cells-substituted sustains P)
- (R4) (N-with-2%-cells-substituted) sustains P
- (R5) (N-with-3%-cells-substituted) sustains P
-
- (R*) (N-with-100%-cells-substituted) sustains P (Goodenough 1996: 114)

that the conjunction of (i) and (ii) commits Williams to the implausible claim there is a sharp dividing line in the Physical Spectrum.

As stated, this is certainly a sorites argument (of the sort I will discuss in more detail in Chapter 2). We begin by acknowledging, in effect, that the “near” operation in the Physical Spectrum “sustains” my identity, *i.e.* that I will survive an operation in which nothing is done to my brain and body at all. Then, in (R2) and (R3), we are supposed to agree that, if I would survive an operation O , then I would survive an operation which differs from O only in the substitution of an extra hundredth of my cells. *Via* orthodox logic (universal instantiation and *modus ponens*), we are to be led to the conclusion (R*) that even were my entire brain and body to be physically destroyed and replaced with new but qualitatively identical matter, I should still survive the operation – my personal identity would still be “sustained”. This conclusion is inconsistent with the Physical Criterion. That, then, is how Goodenough (1996: 114) thinks that Parfit argues against the Physical Criterion.

1.5.4 Goodenough’s critique of the Replacement Argument

Having thus interpreted Parfit, Goodenough goes on to argue that anyone who accepts the Replacement Argument is committed to certain conclusions which should be unacceptable to anyone who, like Parfit, ‘cleaves to a broadly materialist view of the universe and its contents’. (Goodenough 1996: 117) Goodenough argues that Parfit’s alleged commitment to

(R2) If N sustains P , then (N -with-1%-cells-substituted) sustains P .

and to

(R3) If (N -with- x %-cells-substituted sustains P) then (N -with- $x+1$ %-cells-substituted sustains P).

also commits him to

(S2) N -1%-cells sustains P .

and to

(S3) If (N-1%-cells sustains P) then (N-(x+1)% cells sustains P) [.]

(Goodenough 1996: 117)

Thus, argues Goodenough, Parfit is committed to a sorites argument from (R1) ('N sustains P'), (S2) and (S3) to the following conclusion:

(S*) (N-100%-cells) sustains P [.]

(Goodenough 1996: 117)

(S*) says that a person would continue to exist even after the complete annihilation of his/her brain and body. If Goodenough is right, then, Parfit is committed to a highly contentious proposition, which should be unacceptable to anyone who, like Parfit, 'cleaves to a broadly materialist view of the universe and its contents'.

1.5.5 Goodenough misinterprets Parfit

I do not wish to dispute Goodenough's claim that *if* Parfit is committed to the Replacement Argument, *then* he ends up having to defend (S*). But I do dispute his interpretation of Parfit as endorsing the Replacement Argument. Parfit in fact neither endorses, nor commits himself implicitly to, the Replacement Argument – even though, in his rejecting the Physical Criterion, he endorses its conclusion.

Let us address, then, the exegetical question of whether Parfit consciously endorses the Replacement Argument. Perhaps on a quick reading, one might come to the conclusion that he does. Having described the Physical Spectrum, Parfit asks:

If they replace only 1%, would I cease to exist? This is not plausible ... But what about the cases where they would replace 5%, or 10%, or 30%, or 60%? (Parfit 1984: 235)

It sounds here as if Parfit is challenging the defender of the Physical Criterion to locate an exact point of transition in the Spectrum between those cases the patient determinately survives, and those s/he does not. Says Parfit, 'This range of cases challenges the Physical Criterion.' (1984: 235)

However, Parfit shortly afterwards goes on to suggest a way out for the defender of the Physical Criterion. He suggests that the Physical Criterion may be defended by claiming that in operations around the middle of the Physical Spectrum, ‘the question ‘Am I about to die?’ has no answer’. (1984: 235) If there is a range of indeterminate cases in the Spectrum, then the sorites chain is blocked. There may be no *particular* point at which the transition occurs; but this is not to say that there is no transition. The defender of the Physical Criterion may justifiably hold the transition to be gradual, taking place *via* a zone of intermediate cases. In such cases, says Parfit, ‘I do not know whether the resulting person will be me ... [but] this is not here a real question ... [but] an empty question’. (1984: 235) This is not what Parfit *himself* would say about the Physical Spectrum, for he defends a psychologically-based account of DPI which holds that in all operations in the Physical Spectrum identity is maintained (since psychological properties and relations are maintained). However, Parfit explicitly suggests that ‘for those who accept the Physical Criterion, this is the right reaction to this range of cases.’ (1984: 235) This amounts to a rejection by Parfit of the Replacement Argument as an effective argument against the Physical Criterion. In other words, Parfit rejects the idea that a criterion of personal identity ought to be rejected merely because, in a suitably constructed sorites series, there would be problems using that criterion to determine a sharp boundary. Not only would this be inconsistent with Parfit’s comments just mentioned (1984: 235), but this would also be inconsistent with Parfit’s insistence elsewhere that questions of personal identity may be ‘empty’, and have no right or wrong answers. (*E.g.* Parfit 1984: 217, 239.)

In summary, Parfit does *not* endorse the Replacement Argument. He does not think this a good argument against the Physical Criterion.

The force of Parfit’s Physical Spectrum is as a challenge to one who, like Williams (1970), would *both* affirm the Physical Criterion *and* deny Personal Indeterminism; for one would then have, implausibly, to posit a *sharp* boundary in the Physical Spectrum, in order to block the Replacement Argument.

1.5.6

Goodenough would reply by saying that, even if Parfit does not explicitly endorse the Replacement Argument, he nevertheless commits himself to its premisses. Says Goodenough:

Either [Parfit] must countenance some dividing line in the proceedings in order to block the inference chain (S4), (S5) ... (S*), which will negate the whole point of his [s]orites reasoning against dividing-lines in these cases, or he must be prepared to accept the consequences of a possible non-physical personhood ... (Goodenough 1996: 119)

Goodenough's line of reasoning here appears to be as follows. (Note: (S4) and (S5) are analogues of (R4) and (R5).)

Given that Parfit denies sharp boundaries in such series as the Combined Spectrum, it might be argued that he is committed in any case to (S2) and (S3):

(S2) N-1%-cells sustains P.

(S3) If (N-1%-cells sustains P) then (N-(x+1)% cells sustains P)

(Goodenough 1996: 117)

For if Parfit denies sharp boundaries, then, one might argue, he denies that there is any K that is a boundary, *i.e.* any K for which K-cells sustains P but K-1%-cells fails to sustain P. And, one might argue, if there is no such K, there is no point in the series that would falsify (S3); and Parfit would then be committed to the argument to (S*) after all:

(S*) (N-100%-cells) sustains P

(Goodenough 1996: 117)

Does Parfit need to be worried by such a rejoinder? No. There is at least one plausible way of denying there is a *sharp* boundary without being committed to S3. One may hold, as under supervenience (see Subsection 2.2.3), that there *is* a K for which K-cells sustains P but K-1%-cells fails to sustain P, yet still maintain that the facts do not determine just *which* point in the series K is. Thus, one may still deny that there is a *sharp* boundary, by denying that there is a (unique) boundary with determinate location. This proposal will receive further justification in Chapter 2, especially Section 2.5.

Parfit words his denials of sharp boundaries somewhat loosely. For example, in describing what he believes to be an *implausible* claim about the Physical Spectrum, he says:

[It might be claimed that] [t]here must be some critical percentage which is such that, if the surgeons replace less than this percent, it would be me who wakes up, but if they replace more than this percent, it would *not* be me. (Parfit 1984: 235)

A superficial reading would suggest that Parfit, in rejecting the postulation of such a ‘critical percentage’, would commit himself to the claim there is *no* ‘percent’ dividing the cases where I survive from the cases where I do not. And to claim literally this is to be committed to (S3), at least on classical logic, and on some non-classical logics too.

However, it appears most reasonable to interpret Parfit as trying here to express just the claim that there is no *sharp* boundary – and, as I will argue in Chapter 2, this claim (even given classical logic) does not commit one to such “tolerance principles” as S3.

Parfit hints at the necessary further qualification when he uses, for example, the adjective ‘critical’ in the above passage; or when he finds it ‘hard to believe that, in one of [the cases in the Combined Spectrum], the resulting person would quite *straightforwardly* be me [*P*], and that, in the next case, he would quite *straightforwardly* be someone else.’ (1984: 239) (My emphasis)

Let us turn now to a more substantial challenge to Parfit than Goodenough’s. This comes from Bernard Williams (1970) – from a different part of the same paper in response to which Parfit framed the Physical Spectrum. Williams argues that there is something special about *personal* diachronic identity, such as to exempt SDPIs from the possibility of indeterminacy.

1.6 WILLIAMS’S TORTURE CASE

1.6.1

In Section 1.4, it was argued that the problem of locating a sharp boundary in the Combined Spectrum is just as great when considered from a first-person perspective,

as when considered from a third-person perspective. On account of the gradualness of psychological difference across the Spectrum, it is likely that there would be some cases where the patient, after the operation, is unable to say confidently either that s/he is identical to *P*, or that s/he is not identical to *P*. From the point of view of the post-operation patient in such a case, there is an unclarity associated with the exercise of looking into the past, and asking, 'Am I identical with the pre-operation person?' The plausibility of supposing that there would be this unclarity lent some support to the case for Personal Indeterminism, because it lent support to Parfit's *second point*, about the difficulty of locating a sharp boundary in the Combined Spectrum.

1.6.2 Anticipating one's future

So much for the exercise of looking, from a first-person perspective, into one's past. But what about the exercise of looking into one's *future*? A person, unlike a building, generally can *anticipate*, from a first-person perspective, what his/her future will be like for him/her. Bernard Williams (1970: 174-180) has argued that this possibility of anticipation renders problematic the notion of indeterminate SDPIs. (See also Swinburne 1974 & Swinburne 1984: 17-19.)

What I have called 'anticipation', Williams calls 'projective imaginative thinking'. (1970: 177) This refers to the exercise of imagining "what a certain future situation will be like for me", or in other words, "what the future will be like for me", or in other words "what it will be like for me".

Suppose I know that tomorrow I will be involved in a situation constituted by my playing pool with three other people. I might imagine this situation from a perspective that is neutral as between the four persons involved. I do this by imagining four people standing around a pool table. In my imaginative picture here I might, say, imagine what the situation would look like from above. That one of the people happens to be me does not affect in any special way my picture of this situation. The person who is pictured here as *me* is simply someone who happens to look a certain way, to wear certain clothes *etc.*. Hence, I identify, within this picture, who of the four is *me* only by this person's "outwardly visible" characteristics as exhibited within the picture. This person is not privileged by my picture of the situation.

But instead of imagining this situation from an impersonal perspective, I might rather *anticipate* what this situation will be like *for me*. This involves imagining the

very same situation, but in a way that privileges one of the four pool players. The privileging of this player is not a feature of the situation itself. It is a feature of my picture of this situation. This feature involves, among other things, the fact that, in this picture, there is one person whose head is never seen (unless in a mirror). It also involves the fact that, whenever this person is imagined taking a shot, this person's cue appears in the picture from a certain peculiar angle; and a certain sensation is present in the picture, being the sensation of striking a ball with a cue; whereas when others take a shot, their cue is never seen from that angle, and the sensation of taking a shot is not included in the picture. To anticipate what some situation will be like *for me*, I choose one person whose perspective within that situation I will privilege in this way; and I do not correctly *anticipate* this situation unless the person whom I thus privilege *is* me.

1.6.3 The Torture Case

Williams (1970) poses the following hypothetical scenario, which I will call 'Williams's Torture Case'.¹² Suppose I am told that tomorrow, a person *X* will be tortured, and that it is indeterminate whether I am diachronically identical to *X*. Thus, it is indeterminate whether I will be tortured tomorrow. It is determinately the case that someone will be *tortured*. But, I am told, it is indeterminate whether the person to be tortured will be me. This indeterminacy will not be removed when the torture occurs, as it would had I been told that either myself or a certain other person will be randomly selected to be tortured. Rather, the indeterminacy will remain even after the torture has occurred.

Williams argues that I would be faced with two peculiar difficulties were I given this information.

1.6.4 The difficulty of projective imaginative thinking

The first difficulty is the difficulty of anticipating what it will be like for me tomorrow. Williams argues that if I engage in 'projective imaginative thinking (about how it will be for [me])', *i.e.* if I anticipate what tomorrow will be like for me, then I will be answering 'the necessarily unanswerable question'. (Williams 1970: 177) But

‘if [I think I] cannot engage in such thinking, it looks very much as if [I] also [answer] it, though in the opposite direction.’ (1970: 177) Williams suggests that I might refrain from projective imaginative thinking altogether. However, he queries whether this would really be ‘*just* refraining from it, if it is incurably undecidable whether [I] can or cannot engage in it’. (1970: 177) (My emphasis.)

Williams’s point might be re-cast as follows. I know that a certain situation, involving the torture of someone such that it is indeterminate whether that someone is me, will occur tomorrow. I might imagine this future situation from an impersonal perspective. I might imagine, “from above”, some human being (the “torturee”) being tortured. But if I try to *anticipate* what tomorrow will be like *for me*, I have two options. I might privilege the perspective of the torturee, *i.e.* imagine the situation from the torturee’s point of view. But my attempted anticipation of my future is a correct anticipation of what my future will be like for me only if the future person whose perspective I privilege is diachronically identical to me; and in this case, it is not determinate that the torturee *is* identical to me. The other option is to imagine the situation without privileging the perspective of the torturee. But again, this would not be correctly to anticipate what tomorrow will be like for me, because it is not determinate that I am *not* the torturee. It is thus impossible for me correctly to anticipate what this situation will be like for me. Williams suggests that perhaps it would even be inappropriate for me to *refrain* altogether from attempting to anticipate what this situation will be like for me. The suggestion here appears to be that, by refraining from imagining the situation in that anticipatory way, I incorrectly characterise it as one in which I am determinately not involved – whereas it is *indeterminate* whether I am involved.

1.6.5 The difficulty of responding emotionally

In addition to this difficulty of ‘projective imaginative thinking’, Williams also argues that I would be faced with another peculiar difficulty, *viz.* that of forming an appropriate *emotional* response to the information. (1970: 177-178) Should I *fear* the torturee’s torture? If I adopt an attitude of fear (in the sense of “anticipatory dread” or “fear of undergoing”), I am adopting an attitude that is appropriate only in case the

¹² Williams’s Torture Case is actually an extraction from a larger hypothetical scenario of which his (1970) paper is a discussion; but it can be treated as a separate hypothetical case in its own right. I shall (footnote continued next page)

torturee is diachronically identical to me. But it is indeterminate whether the torturee is identical to me. Conversely, if I adopt an emotion that is not fear, but rather sympathy or the like (or perhaps “fear” in the sense of “fear for” someone else), then I am adopting an attitude that is appropriate only if the torturee is not identical to me. But it is indeterminate whether the torturee is not identical to me. Thus, as well as the “imaginative” difficulty discussed above, I would also be faced with an “emotive” conundrum.

Williams emphasises that my emotive difficulty here is more radical than the difficulty I would face were I told that some inanimate object I hold dear will be tampered with in such a way that it will be indeterminate whether the product of the tampering is diachronically identical to the object which I now hold dear. (Imagine a teddy-bear, for example.) In that case, I might have some intermediate or ambivalent emotion towards the “transmogrified” object. I would not be as attached to it as I was to the original object as it was pre-transmogrification; but neither will I be completely indifferent towards it. I might be faced with some kind of emotional strangeness here; but Williams emphasises that it would be nothing like my situation in the torture case. In the torture case, it would be incorrect to regard the torturee as one would regard the transmogrified bear: that would be to regard him/her from a third-person perspective, and thus to assume that I will *not* be the torturee. I would thus ‘displace the conceptual shadow from its proper place’. Williams argues: ‘I have to get nearer to [the torturee] than that. But is there any nearer that I can get to him without expecting his pain?’ (Williams 1970: 177)

1.6.6 Seeking arbitration

Finally, Williams argues that to propose to resolve my conceptual difficulty by *arbitrating* whether I shall be identical to the torturee, would be completely inappropriate, and would not ease my difficulty at all. (Williams 1970: 178) We might illustrate Williams’s point here by comparing the Torture Case with the case of the Santa Trinita bridge. (See Subsection 1.3.2.) Suppose that some practical matter hinges on the question of whether the current bridge at Santa Trinita has existed for at least two hundred years. For example, a travel writer might promise to local Santa Trinita authorities that she will make mention, in her forthcoming book about old

discuss the larger scenario – Williams’s “twin thought experiments” – in Chapter 3.

bridges, of any of Santa Trinita's bridges that are at least two hundred years old. If she then omits from her book the bridge at Santa Trinita, the question of whether she has breached her promise might reasonably be resolved by arbitration by a magistrate. However, in Williams's Torture Case, it seems that arbitration about whether the torturee is identical to me would do nothing whatsoever to resolve my conceptual and emotive difficulties. As Noonan (1989: 191-192) points out in his discussion of the Torture Case, were I to be told before tomorrow that everyone has agreed, from tomorrow and thereafter, to use the word 'person' in such a way that I am regarded as determinately "not the same person as" the torturee, this would appear to give me no reason whatever to be relieved.

1.6.7 Noonan on the Torture Case

Should Williams's Torture Case convince us that diachronic personal identity statements cannot be indeterminate? Noonan (1989: 190-195) considers this question, and answers it in the negative. In trying to anticipate what tomorrow will be like for me, and in trying to form an appropriate emotional response, it seems that I really would face peculiar difficulties. Noonan suggests that an opponent of Personal Indeterminism might contend that 'my difficulty has a straightforward explanation', viz. 'that what I have been told to expect is impossible'. (1989: 191) But clearly this is not the only explanation, and if we have strong independent grounds for being Personal Indeterminists – e.g. the Combined Spectrum Argument – then it is an explanation that we should reject. Our inability to envisage a situation in a certain way simply does not entail that this situation is impossible. It could be that the situation is possible, but we just do not have the capacity to think about the situation in a certain way, or to form an adequate, fitting or coherent emotional response to it.

Noonan suggests that, in the Torture Case, I might respond by maintaining that I 'cannot engage in ... projective imaginative thinking about how it will be for [me]', while 'giv[ing] as [my] reason that to do so would be implicitly to answer the necessarily unanswerable question'. (1989: 192) (Noonan's emphasis.) As Noonan points out, it is not clear that *this* way of responding would commit me to the view that the torturee determinately is not me. And it certainly does not commit me to the view that the torturee determinately *is* me.

Noonan goes on to point out that even if in the Torture Case I cannot correctly *anticipate* what tomorrow will be like for me, it does not follow that I cannot make

any adequate representation of this situation at all. (Noonan 1989: 192-193) I might picture the entire scenario from an impersonal perspective. I might imagine some kind of brain-altering and/or body-altering operation being performed on a certain human being, such that the person on the operating table after this operation is neither determinately identical to, nor determinately non-identical to, the person on the table before the operation. I might then imagine that the person who emerges from this operation is tortured. I might imagine all of this from an impersonal (third-person) perspective.

But even if I could not imagine the scenario in *any* way, it still would not follow that these events would be impossible. It might simply be that my imaginative capacities are limited.

Similar points can be made with regard to the difficulty of forming an appropriate emotional response. It does not follow from this difficulty that the situation to which I am trying to respond is itself impossible or incoherent. As Noonan points out, if I have not been confronted with a particular situation before, it may be unsurprising that I am unable to form an appropriate emotional response to it. (1989: 193) Williams's argument assumes 'that one must be able to produce an appropriate emotional response to any genuinely possible situation which one can foresee.' (Noonan 1989: 193) But there is no reason to assume this.

With regard to the thought that one might seek arbitration about the question of one's identity, Noonan agrees with Williams that 'it would be absurd to let one's emotional response ... be determined by ... linguistic refinements or revisions the people around one choose to agree on to enable them to pigeon-hole the situation in the way they find most convenient.' But this does not entail that the situations themselves are not 'conceptually possible'. (Noonan 1989: 192) (I shall revisit this matter in Section 3.7.)

1.6.8 There is no "intermediate perspective"

Williams's Torture Case, then, should not persuade us to reject Personal Indeterminism. However, it does bring to light an interesting aspect of how we tend to conceptualise our own futures.

Recall the situation of four people playing pool. In imagining this as a future situation, I may adopt an impersonal perspective, privileging none of the four pool players. Or I may imagine the situation in a way that privileges one or another of the

players, in the way described earlier. (The privileged person's head is unseen within the picture, *etc.*) In other words, I assign the first-person perspective to one person, and assign the third-person perspective to the three others.

Call the four pool players 'A', 'B', 'C' and 'D'. Suppose it is indeterminate whether A is identical to me, or B is identical to me, while it is determinate that neither C nor D is identical to me. If I am correctly to anticipate what this situation will be like for me, I must assign the third-person perspective both to C and to D. But what perspective should I assign to A? And what perspective should I assign to B? If there were a perspective that is intermediate between the first-person perspective and the third-person perspective, perhaps that would be the correct perspective to assign both to A and to B. But there is no such "intermediate perspective". Furthermore it is impossible, or at least rather challenging, to imagine this situation in such a way as to assign the first-person perspective to *both* A and B. And if I assign the first-person perspective to no-one, this amounts to envisaging the situation impersonally, rather than anticipating what it will be like for me. Williams's Torture Case brings to light that our imaginative capabilities are limited in *these* ways – even though, in other ways, for example from an impersonal perspective, we seem coherently to be able to represent situations of indeterminate DPI. (Consider the central cases in the Combined Spectrum.)

The temptation to deny that SDPIs can be indeterminate seems to stem, at least in part, from the inability to adopt an intermediate perspective.

1.6.9 More on anticipation

Let us continue with the above example. It is indeterminate whether A is identical to me. Now consider two mental exercises. In the first I imagine what the situation will be like for C (with whom I am determinately non-identical). Do we call this exercise an 'anticipation'? No. Why? The answer is 'No' *simply* because it is determinate that I (the imaginer) am not identical with C. Now consider a situation in which I imagine what the future will be like for A. Does this count as an anticipation? The apparent mystery with which Williams was grappling was merely a "mystery" about what to say about whether such an imaginative exercise counts as 'anticipation'. But, analytically, it counts as *anticipation* if and only if A is identical with me (the imaginer). So since it is indeterminate whether A is me, it is indeterminate whether the mental exercise of imagining what the situation will be like for A is anticipation.

Where x imagines what the future will be like for y , x engages in an exercise of *anticipation* if and only if $x=y$. If $x\neq y$, x is said to engage in empathic thinking, or something like that.

For me to imagine what tomorrow will be like *for* y is simply for me to form an imaginative picture of the world, but from a certain pictorial perspective. (A picture of a train might yet depict the wheels, but not the inside of the engine. It might depict one side of the train, but not the other.) Depending on the perspective – which is, of course, a feature of the picture and not of the world – we say that this picture represents what it is like for y , or for z , or *etc.*. If the person who is doing the imagining is diachronically identical to the person whose perspective is privileged in the picture, then we say that the imaginer, in forming the picture, engages in anticipation.

One might ask: But is there not more to the concept of anticipation? Is it not rational for me to have a “special concern” about a certain future person, because that person is me? (See Parfit (1984: 307ff.) for the notion of “special concern”.)

What would it be to be especially concerned about a person x ? Presumably it would involve, among other things, thinking about x a lot, attending to x 's needs, and frequently imagining what certain situations will be like for x . It seems rational to be especially concerned about one's particular loved ones in *these* ways; and it seems no less rational to be especially so concerned about oneself. (See Johnston 1992: 597-600, and Wolf 1986, esp. pp. 718-720.) While the question of how a rational person would structure his/her concerns about persons existent in the future has subtle aspects which I will not explore, the point I want to make here is simply that, whatever special concern I do have for my future self, in comparison with my concern for future others, seems to be special in *degree*, and not in *kind*. There is no really distinctive type of psychological exercise I undertake in imagining what the future will be like for me. I simply form an imaginative picture of the world (which I happen to call an ‘anticipation’, because of certain perspectival features of the picture), of essentially the same sort as that which I would form in imagining what the future will be like for someone else. I do imagine what the future will be like for me more *often* than I imagine what it will be like for others. And I am better at predicting what tomorrow will be like for me than at predicting what it will be like for others, because I have *more* information about the factors that influence my future than about the factors influencing the futures of others. But these are matters of degree, not of kind.

Suppose that, over the duration of today, I imagine what tomorrow will be like for me, and then what the next day will be like for me, and then the next, and so on. Eventually I get to the last day of my life. I imagine for all the events up until my death what they will be like for me. And then at this point it feels like there is a mystery. I want to ask: What will it be like for me after I am dead?

Assuming that I do not exist after my death, we can answer this question as follows. No imaginative exercise in which I generate a picture of the world after my death counts as anticipation, or as ‘imagining what it will be like for me’, simply because no such picture in which events *after* my death are depicted could privilege the perspective of a person x such that $x = \text{me}$, since in no such events is there a person x such that $x = \text{me}$. But this follows trivially from the proposition that I do not exist after my death.

Nevertheless, I can still imaginatively depict what the *world* will be like after my death. And depicting the *world* is all I have *ever* done. Some imaginative pictures of the world count as anticipations; others do not. To discriminate between the former pictures and the latter simply involves applying the criteria of diachronic personal identity (be they physical, psychological or *etc.*), with reference (a) to the present person doing the imagining and (b) to the person perspectively privileged within the picture. (“Simple Theorists” of DPI may disagree, since they deny that there are any criteria of DPI; I critique that view in Chapter 3.)

I shall return to the notion of anticipation in Chapter 3, in response to a certain “broader argument” for the Simple View of DPI.¹³

1.7 A SEMANTIC FRAMEWORK

1.7.1

Assuming that a statement of (personal or non-personal) diachronic identity $\langle x, y, t, t^* \rangle$ is indeterminate, what is the *source* of this indeterminacy? I propose in this section to distinguish four answers to this question. These answers are not theories about the correct logical treatment of indeterminacy: such theories will be discussed

¹³ For a different take on the notion of anticipation, see Martin (1995).

in Chapter 2. These answers presuppose, furthermore, that there *can* be indeterminacy – a presupposition which I will test in Chapter 2. Rather, these answers say something about whence the indeterminacy of $\langle x, y, t, t^* \rangle$ (whatever its nature) *arises*, assuming it does arise.

Given that $\langle x, y, t, t^* \rangle$ is a statement of diachronic identity, it can be assumed that x does exist at t and y does exist at t^* . So if $\langle x, y, t, t^* \rangle$ is indeterminate, it is indeterminate whether $x=y$.

If ' $x=y$ ' is indeterminate, there are four (not obviously mutually exclusive) things that might be said about *why* it is indeterminate. The indeterminacy might be sourced in:

- (1G) indeterminacy-generating features of either or both of the singular terms ' x ' and ' y ';
- (1H) indeterminacy-generating features of the predicate '=';
- (1I) indeterminacy-generating features of the object(s) x and y ;
- (1J) indeterminacy-generating features of the identity relation (as opposed to the identity predicate).¹⁴

I propose to assume, from Chapter 2 on, for the sake of avoiding certain complications, that only 1G is the correct account of why ' $x=y$ ' is indeterminate. I do not propose to defend this assumption at any length, but presently I will give some *prima facie* reasons for rejecting each of 1H, 1I and 1J; this will serve also to distinguish these accounts and to clarify the nature of what 1G is asserting to be true. Further, I will outline a particular account of *how* 1G is true, which account seems to be the only way of making sense of 1G, while rejecting 1H, 1I and 1J.

¹⁴ One might add that perhaps a statement can be indeterminate because of the way its elements are syntactically put together. I will ignore that possibility here.

1.7.2 Ontic indeterminacy

To say that indeterminacy is ontic is to say that it is not merely semantic – it does not arise merely from features of language – but is, rather, a feature of the world itself, had independently of language. A *disbeliever* in ontic indeterminacy (who believes in indeterminacy at all) will source the indeterminacy of an indeterminate statement in features – typically *vagueness* – of linguistic items in the statement, especially of *predicates* and *singular terms*. S/he will thus, in the case of indeterminate diachronic identity statements, favour account 1G or 1H.

In contrast, a believer in ontic indeterminacy holds that where a statement is indeterminate, this indeterminacy is not (or is not always) just the result of indeterminacy-generating features of the predicates and singular terms in the statement, but is the result rather of indeterminacy-generating features of the *properties, relations* and *objects* denoted by these predicates and singular terms. S/he holds, then, that the *world* is vague, or indeterminate, independently of language. If there is ontic indeterminacy, then it may be that some object is the borderline case of some vague *property*; or it may be that some predicate is such as to be neither determinately satisfied, not determinately not satisfied, by some *fuzzy* or *vague object* – hence generating indeterminacy only *derivatively* in the resultant statement, but *primarily* in the ontic *state of affairs* that the statement describes. That is, if there is ontic indeterminacy, it may be indeterminate, independently of language, whether a given state of affairs holds, even if we can precisely *describe* this state of (fuzzy) affairs. Clearly, an ontic account of the indeterminacy of a diachronic identity statement would involve defending account 1I or 1J.

Many people do not think there can be any ontic indeterminacy at all. And many have argued that, at least, the *identity relation* cannot be indeterminate. Those arguing this have generally admitted that identity *statements* can be indeterminate. But they source the indeterminacy of these statements, usually, in the vagueness of one or both of the singular terms flanking the identity predicate. For a vague singular term, there is no one object to which it definitely, uniquely refers. It is not the case, for those offering this semantic explanation, that an indeterminate identity statement has both of its singular terms denoting precisely; for then, assuming the numerical identity *predicate* is not vague, any indeterminacy would have to arise from an indeterminacy-generating feature in the identity *relation*, or from such a feature in the objects denoted; but this would be ontic indeterminacy.

The basic argument against indeterminacy regarding the identity relation is this. It was proposed independently by Evans (1978) and Salmon (1981: 244ff). Suppose it is indeterminate whether $x=y$. If the identity *relation* can be indeterminate, then x could have the *property* of being such that it is indeterminate whether it is identical with y ; but since this is a property not shared by y , $x \neq y$. That we can *derive* that $x \neq y$ seems to rule out its being indeterminate whether $x=y$. But surely it *can* be indeterminate whether $x=y$. (Consider, *e.g.*, the Inanimate Spectrum.) Therefore there is no such *property* as ‘being indeterminate whether one is identical with y ’. Therefore the identity *relation* cannot be indeterminate. To say that it is indeterminate whether $x=y$ is not to say something about a relation between a (precisely denoted) object x and a (precisely denoted) object y . Rather, it is to say something about relations between linguistic expressions in ‘ $x=y$ ’, and the world. That is the only kind of explanation we can give of the indeterminacy of ‘ $x=y$ ’ which does not entail that x has some property that y lacks; and since x has a property lacked by y only if determinately $x \neq y$, it is the only kind of explanation we can give of it at all. (But because this explanation *is* available, we need not conclude that there can be no indeterminate identity *statements* at all – as I alluded in Subsection 1.3.5.)

This argument has of course been contested: see, *e.g.*, Lowe (1994), Keefe (1995) and Parsons and Woodruff (1995). Furthermore, even if we accept the argument, perhaps we can still maintain that *objects* (though not the identity relation) can have indeterminacy-generating properties, such as “fuzzy-boundedness”: Garrett (1998: 80) argues that the Evans-Salmon argument does not establish that objects cannot be like this.¹⁵

I propose to avoid the intricacies of the debate about ontic indeterminacy so that my own investigations will be able to proceed more smoothly. I therefore make the unargued (or very under-argued) assumption that indeterminacy is never ontic, *i.e.* that 1I and 1J are incorrect. I take the anti-ontic option for two reasons.

Firstly, I prefer that option – perhaps more due to nebulous intuitions than because I am persuaded by the Evans-Salmon argument (though I find it hard to fault). It seems to me that almost all the (comparatively) uncontroversially indeterminate statements actually uttered (statements of the ‘Fred is bald’ type)

¹⁵ For defences of and ramified versions of the Evans-Salmon argument see Noonan (1989: 112-117) (the second edition – significantly updated here) and Garrett (1998: 73-82).

contain words which it is very natural to call vague, and to the vagueness of which it just seems natural to trace the indeterminacy of the statement; and it seems to me that the indeterminacy of indeterminate identity statements can plausibly be traced to the vagueness of words too. (See Subsection 1.7.4.) So an appeal to ontic indeterminacy seems to me unmotivated with respect to such statements. Perhaps there is ontic indeterminacy at the level of tiny particles (see Lowe 1994, Chibeni 2004), even if there is not ontic indeterminacy at the level of persons and tables; but then I do not find this possibility particularly relevant to my interest in indeterminate statements of diachronic personal identity.

Secondly, the semantic account of indeterminacy seems the more orthodox one.

1.7.3 The identity predicate

Let us now consider 1H.

The numerical identity predicate seems pretty precise. To say that x is numerically identical with y is to say (at least) that there is *just one* object in question. There might be disputes about identity statements, but these would seem to revolve around the singular terms: whether they are proper names, definite descriptions, rigid designators *etc.*. But these disputes would seem not to impinge on the core of the numerical identity concept: the notion that there is just *one* object in question. As long as this core is undisputed, there seems little room for indeterminacy, it would seem, in the predicate '=' itself.

Sider (*e.g.* 1996) has analysed diachronic identity statements, or at least *tensed* ones, in terms of a relation of temporal counterpartcy between numerically distinct temporal object-stages, rather than in terms of strict numerical identity between a perduring aggregate of object-stages and itself, or between an enduring object and itself.¹⁶ For Sider, a currently existent person *did* exist (or *will* exist) at a certain time if and only if s/he has a temporal counterpart at that time. He has suggested, of at least some cases where it is indeterminate whether person P is diachronically identical with Q , that the indeterminacy arises from there being multiple candidate counterpartcy relations. (2001: 193) So we might say that where $\langle x, y, t, t^* \rangle$ is indeterminate, the

¹⁶ See Noonan (1989: 100-104) for one explanation of the difference between perdurance and endurance. See Note 18, below, for my brief explanation of this difference.

Siderian account sources the indeterminacy in a property of the diachronic identity predicate, *viz.* its ranging over multiple candidate counterpartcy relations.

Sider's ontology of persistence is unorthodox, and though I have nothing against it, I will assume it is false for the sake of avoiding having to make too many provisos in discussing diachronic identity statements. That is, I will assume the diachronic identity predicate denotes unequivocally the numerical identity relation. But I doubt Sider's ontology would give him a *particular* reason to reject appropriately re-worded versions of my assertions.

An aside: If it were indeterminate whether 'is identical to', in the context of a diachronic identity statement, denotes numerical identity or temporal counterpartcy, *that* indeterminacy could hardly be the source of the indeterminacy of indeterminate diachronic identity *statements*. That is, it is hard to see how, if $\langle x, y, t, t^* \rangle$ is indeterminate, so that it is indeterminate whether x is y , its indeterminacy would dissolve were 'is' made either unequivocally Siderian or unequivocally non-Siderian.

I will assume, henceforth, that 1H is incorrect.

1.7.4 The object-candidate account

1G would account for the indeterminacy of ' $x=y$ ' by reference to indeterminacy-generating properties of at least one of the singular terms ' x ' and ' y '. I will assume that 1G is correct. And I suggest that the following "object-candidate account" is the best way to make sense of 1G. This account is drawn directly from Noonan (1989: 108-112).

Consider Shoemaker's (1984: 145-6 n. 5) example of Alpha and Beta Hall, two structures joined by a narrow walkway. Suppose it is indeterminate whether they together constitute one building, or two. Suppose Smith is lecturing in Alpha Hall and Jones is lecturing in Beta Hall. Let ' a ' abbreviate 'the building Smith is lecturing in', and let ' b ' abbreviate 'the building Jones is lecturing in'. So it is indeterminate whether $a=b$.

In saying 'the building Smith is lecturing in', one creates a constraint: whatever ' a ' denotes, it denotes a building. But because of the vagueness of 'building', it is indeterminate whether the entire structure (consisting of the two halls and the walkway joining them) counts as a building, or whether each hall counts as a distinct building. So, given that ' a ' denotes a building, it is indeterminate whether it

denotes Alpha Hall, or the entire structure. The same may be said about 'b', except that its denotation is indeterminate as between the entire structure and Beta Hall.

For each singular term there exist multiple candidate referents. The indeterminacy of ' $a=b$ ' is traced to the fact that, for some pairs $\langle a^*, b^* \rangle$ of candidate referents for 'a' and 'b' respectively, $a^*=b^*$ (viz. the pair \langle the entire structure, the entire structure \rangle), and for other such pairs $\langle a^{**}, b^{**} \rangle$, $a^{**} \neq b^{**}$ (e.g. the pair \langle Alpha Hall, the entire structure \rangle).

Call this the 'object-candidate account' of the indeterminacy of ' $a=b$ '.

We can apply this account to an indeterminate statement of diachronic personal identity. Suppose x is the person existent before, and y is the person existent after, some transmogrifying procedure of the sort argued to occur in some of the central cases in the Combined Spectrum. It is indeterminate whether $x=y$. Then, according to the object-candidate account, there exist the following objects: an object u which exists before but not after the operation; an object v which exists both before and after the operation; and an object w which exists after but not before the operation. It is indeterminate whether 'x' denotes u , or v ; and it is indeterminate whether 'y' denotes v , or w . The indeterminacy of ' $x=y$ ' is thus traced to the existence of these multiple candidate referents for each term such that for some such pair of candidates $\langle x^*, y^* \rangle$ for reference by 'x' and 'y' respectively, $x^*=y^*$ (viz. the pair $\langle v, v \rangle$), and for another such pair $\langle x^{**}, y^{**} \rangle$, $x^{**} \neq y^{**}$ (viz. the pair $\langle u, w \rangle$).¹⁷

1.7.5 A challenge for personal endurantists

Is there anything to dislike about the object-candidate account?

One thought is that personal perdurantism can better accommodate the object-candidate account than personal endurantism.¹⁸ This might be thought a problem for

¹⁷ I call this the 'object candidate account' because the candidates are objects. Alternatively, the candidates might be something like these: $\{x: x$ satisfies the definite description ' u ' $\}$, $\{x: x$ satisfies the definite description ' v ' $\}$ and $\{x: x$ satisfies the definite description ' w ' $\}$. (They would not then be candidates for *reference* by the singular terms, but for some analogue of reference, perhaps.) This kind of "set-candidate" account would appear to be neutral as between ontic and non-ontic accounts; the former would allow that it is indeterminate whether $\{x: x$ fits the definite description ' v ' $\}$ is empty. (I am not sure whether this account, all told, would be coherent, but it is a direction in which one might try to go.) Another alternative is to treat the candidates as the (precise) *terms* ' u ', ' v ' and ' w ' (rather than objects u , v and w). The candidates would then be candidates for precisification, rather than for reference. (See Chapter 2.)

¹⁸ Personal perdurantism is the thesis that persons exist at multiple times by being four-dimensional summations (mereological unions) of temporal person-parts (or person-stages), some part of which exists at each of those times. Personal endurantism is the thesis that persons exist at multiple times by
(footnote continued next page)

the object-candidate account: it obliges one to commit to perdurantism on pain of ontological awkwardness. Consider the following argument, which I have adapted from Noonan (1989: 110-112, 120-121).

Continuing with the above example, the object-candidate account seems committed to the existence of v , which is an object spatially coincident with x before the operation, spatially coincident with y after the operation, but such that this object is neither determinately a single persisting person, nor determinately an amalgam of two distinct persons, one of which dies in the operation.

So, when asked: 'What *is* v ?' we cannot give any ordinary sort of answer. We cannot assert, ' v is a person', since, on the only plausible version of the object-candidate account, it is *indeterminate* whether v is a person. But then it is mysterious what v is. One offering the object-candidate account is left committed to an entity which looks ontologically problematic.

Noonan's (1989: 110-112, 120-121) solution is to appeal to perdurantism. We can understand the notion of a mereological union of parts of disparate objects. Given the existence of those parts, the existence of the union arguably follows trivially; or at least, the concept of such a union is comprehensible. If objects, including persons, can have temporal parts, there is the union v of temporal person-parts comprising the part (perhaps a proper part, perhaps not) of x that exists only prior to the operation, and the part of y that exists only after the operation. This is our answer to the question 'What is v ?' But the endurantist, who denies that persons have temporal parts, cannot give this answer. So s/he is stuck with the mystery of what kind of thing v is supposed to be.

Noonan does not think (his version of) this argument absolutely conclusive against endurantism. But he thinks that 'given the ontological commitments he must take on board, an opponent of the Determinacy Thesis concerning personal identity [*i.e.* a defender of Personal Indeterminism] has a strong reason for regarding persons as four-dimensional perdurers.' (1989: 121)

Should an endurantist reject the object-candidate account? Only if either s/he is prepared to accept ontic indeterminism, or s/he can think of another way of explicating how features of *singular terms* can generate indeterminacy. The

"enduring", *i.e.* being wholly present at each moment of their existence, rather than existing at each such moment merely in virtue of the existence at that moment of one of their proper parts.

endurantist by definition denies an ontology of person-stages, so s/he seems bound to regard the diachronic identity *predicate* as unequivocally denoting the numerical identity relation had by each enduring object to itself. Sourcing indeterminacy in the equivocation of the diachronic identity predicate between multiple relations between person-stages is not an option for the endurantist, then.

Perhaps the endurantist might think it need not be *determinate* that the problematic object v exists. S/he might say that if v is a person (so that $x=y$), v exists, but if v is not a person (so that $x\neq y$), then v does not exist. But consider Shoemaker's halls. The entire structure exists. Its existence does not hinge on whether it is a building. It might be said that this is because being a building is inessential to the entire structure. Being a building is essential to "the building which is the entire structure". If the entire structure is not a building, there is no such thing as "the building which is the entire structure". But if some object is such that it is indeterminate whether it exists, this seems initially to be an ontic kind of indeterminacy, inherent in the "object" (or "semi-object"). To show that the indeterminacy regarding whether "the building which is the entire structure" exists is *not* a mark of ontic indeterminacy, we must show that there is some object which determinately does exist, which would satisfy the description 'building which is the entire structure' were 'the building which is the entire structure' to be construed as existent, but would not satisfy that description, *yet still exist*, were 'the building which is the entire structure' to be construed as denoting nothing. The indeterminacy is thus located in the description, not the world. So if being a person is essential to v , and it is indeterminate whether v is a person, it is indeterminate whether v exists. But if this indeterminacy is to be demonstrably non-ontic, then being a person must be essential to v merely in the way that being a building is essential to "the building which is the entire structure". Then ' v ' would be a mere definite description, and there would be an object which determinately exists no matter what, which object would satisfy that description were v construed as a building, but would not satisfy it were v not so construed. And *that* object would *not* be determinately a person, and a non-ontic endurantist would have to postulate it.¹⁹

¹⁹ An interesting question, which I will not address here, is: How would the question of a Russellian *versus* a Strawsonian unpacking of the definite descriptions bear on these matters?

There is no plausible way for the endurantist to escape the postulation of some persisting object such that it is indeterminate whether it is a person. His/her best hope for avoiding ontic indeterminism would thus be to show that a non-perduring v would not be such a mysterious thing after all. Some non-mereological analogue of mereological union might be useful for this purpose. Perhaps the union of sets of tensed properties? It seems, at any rate, to be the *endurantist's* problem – I will not be concerned with defending either endurantism or perdurantism here.

1.7.6 Predicate vagueness imported into singular terms

When it is indeterminate whether ‘the person on the table after the operation’ denotes an object existing only after the operation, or an object existing both after and before the operation, the indeterminacy would appear to arise from the descriptive content in the singular term (be it explicit, or imported *via* a prior stipulation that, *e.g.* ‘ y ’ shall denote the person on the table after the operation). It is indeterminate whether (a) v , the object on the table both before and after the operation, is a person, or (b) the object w , the object coming into being only at the operation, is a person. Only one of them is a person: but it is indeterminate which one. Were ‘person’ precise, ‘the person on the table after the operation’ (abbrev. ‘ y ’) would denote determinately, and the SDPI ‘ $x=y$ ’ would not be indeterminate, since there would be only one candidate which would satisfy the definite description ‘ y ’ and only one that would satisfy ‘ x ’ (which imports the description ‘person on the table before the operation’).

The Combined Spectrum is thus rather like the series of dogs considered earlier. Dogs around the middle of this spectrum were borderline cases of ‘is a large dog’. Similarly, where $v(n)$ is the object existing both before and after the n th operation $O(n)$ from the left in the Combined Spectrum, there is some operation $O(k)$ (among many) around the middle of the Spectrum for which it is indeterminate whether $v(k)$ is a person. Thus the operation-straddling objects form a spectrum for the predicate ‘is a person’ in the same way that the spectrum of dogs was a spectrum for the predicate ‘is a large dog’.

We can see then that the Combined Spectrum exhibits that ‘is a person’ is vague, like ‘is a large dog’ is. The (apparent) possibility of indeterminate SDPIs arises from the (apparent) vagueness of ‘is a person’.

If vagueness does *not* give rise to indeterminacy, the Combined Spectrum Argument would fail. In Chapter 2, I shall investigate the epistemic view of vagueness, according to which it does not give rise to indeterminacy.

1.8 CONCLUSION

In Section 1.3, I established that, unless the epistemic view is true, statements of non-personal diachronic identity can be indeterminate. In Section 1.4, I described Parfit's (1984) Combined Spectrum Argument for Personal Indeterminism, the claim that statements of diachronic personal identity can be indeterminate; and I rejected some objections to the argument by Madell (1985). In Section 1.5, I considered a critique of Parfit's reasoning by J. M. Goodenough (1996). If correct, this critique would have undermined the Combined Spectrum Argument. But Goodenough's critique was seen to be mistaken. In Section 1.6, I considered Williams's (1970) Torture Case, and concluded, with Noonan (1989), that this case should not persuade us to reject Personal Indeterminism. Nevertheless, the Torture Case brings to light certain things, such as that our imaginative capacities are limited in certain ways. In Section 1.7, I distinguished various ways of "sourcing" the indeterminacy of diachronic identity statements, and assumed the object-candidate account to be the correct such way, though it poses a challenge for personal endurantists. I reserved judgment, for the time being, about the logic of indeterminacy, and about the question of whether statements can be indeterminate at all. I address these questions in Chapter 2.

I now draw the following conclusion. Personal Indeterminism is true *unless* either (a) the epistemic view of vagueness is correct, or (b) the Simple View (the view "that the DPI facts do not supervene on the non-DPI facts") both is true, and is an effective way of blocking the Combined Spectrum Argument. I will address (b) in Chapter 3. Presently, in Chapter 2, I will consider (a).

Chapter 2

VAGUENESS

2.1 INTRODUCTION

2.1.1

Chapter 1 contained an initial argument for Personal Indeterminism, the claim that statements of diachronic personal identity can be indeterminate. I concluded that we should accept Personal Indeterminism, *unless* either (a) a certain “epistemic view” of vagueness is correct, or (b) the Simple View about DPI both is true, and is a feasible way of resisting the Combined Spectrum Argument. In this chapter I will address (a) whether the epistemic view is correct. Furthermore, I will defend a non-epistemic treatment of vagueness which retains classical logic and bivalent semantics. In the course of this defence, I will offer an account of what *indeterminacy* consists in.

2.1.2 The epistemic view and Determinism

The epistemic view is a view about *vagueness*. It was noted in Subsection 1.2.4 that one *mark* of the vagueness of vague *predicates* is that they can have *borderline cases*: objects for which it is not clear that they satisfy the predicate, and not clear that they fail to satisfy it. This unclarity would seem to remain even if we know all the relevant facts about the object, and despite our being fully competent speakers of the language.

It seems intuitively plausible to suppose that the unclarity surrounding borderline predications arises from their being indeterminate, the relevant facts determining neither that the predicate applies, nor that it fails to apply. This explanation appears also to be reasonable with respect to statements about which we are irremediably unclear on account of their including vague *singular terms*. Consider, for example, ‘I am in Melbourne’, uttered by me when I am in the fringes of Melbourne’s suburbs. The singular term ‘Melbourne’ is vague, at least in certain

A further aim will be to defend a particular understanding of vagueness and indeterminacy, one which requires us to abandon neither classical logic nor bivalent semantics, yet which is not committed to Determinism. But before doing either of these things, I will outline the philosophical debate about vagueness.

2.1.3

An aside: It so happens that the arguments that have been deployed by Determinists generally commit them to the following thesis, which is stronger than Determinism.

(Determinism*) No statement can be indeterminate.

If Determinism* is true then, regardless of whether the Combined Spectrum Argument relies on the assumption that the vagueness of the expressions in a statement can make it indeterminate, Personal Indeterminism is simply false, since it asserts of some statements (*viz.* SDPIs) that they may be indeterminate. Note that being committed to Determinism* does not prevent one from holding that grammatically assertoric utterances in which there is, say, failure of reference, are neither true nor false; for one may hold that such utterances do not say of anything that it is the case, and so do not express statements. Determinists have generally accepted, however, that grammatically assertoric utterances will not fail to express statements simply on account of containing vague expressions. They are thus committed to claiming that an utterance of '*d* is a large dog' (assuming that '*d*' refers) expresses a statement which is not indeterminate. And they ought to be so committed. Should utterances which contain vague expressions be precluded from saying of anything that it is the case, most of our ordinary predicates would be so precluded.

2.1.4

It was noted in Subsection 1.2.4 that one mark of vague predicates is the possibility of borderline cases. To really grasp the philosophical problem of vagueness, however, one must understand another mark of vague predicates: their susceptibility to the construction of sorites paradoxes. I will explain the philosophical problem of the sorites in Section 2.2. It will be seen that the ability to respond satisfactorily to this

²⁰ I do not mean to suggest that Determinism, and the epistemic claim, do not entail each other.

problem serves as a benchmark for theories of vagueness. I will then outline the three approaches to vagueness to have been defended most prominently in recent literature: multivalentism, supervaluationism and, of course, the epistemic view. I will explain how each of these purports to resolve the sorites problem. It will be seen that multivalentism and supervaluationism accommodate, or purport to accommodate, indeterminacy, while the epistemic view does not.

In arguing that we are justified in believing that borderline predications are indeterminate, I will *not* take the approach of defending either multivalentism or supervaluationism. Rather, I will take the following approach. Multivalentism and supervaluationism have in common that they reject bivalent semantics or classical logic (or both). I will ask: If we have good reason to reject multivalentism and supervaluationism, or to reject *any* theory of vagueness that rejects either bivalent semantics or classical logic, must we thereby commit ourselves to Determinism?

I will describe in detail two arguments by Timothy Williamson (1994), which might appear to pose a threat to the Anti-Determinist. Williamson is probably the most discussed recent Determinist, his 1994 constituting a comprehensive case for the position. I will base much of my discussion of the epistemic view on his work. The two arguments are: (1) the “Argument against Supposed Counterexamples to Bivalentism” (an accurate description), which I will refer to more briefly as ‘the Argument for Bivalentism’ (a slightly misleading description); and (2) the “Omniscient Speakers Argument”. Each of these arguments has a conclusion which may seem difficult to reconcile with the denial of Determinism. (1) and (2) are not the only arguments Williamson presents in defence of Determinism, but I think they are the most significant *positive* ones (as opposed to his *negative* arguments, being attacks on specific rival theories, *e.g.* supervaluationism). In Section 2.3, I shall describe the Argument for Bivalentism. I shall outline, without exploring in depth, some lines of defence that may be open to the Non-Bivalentist. But I will not commit myself to any of these lines of defence. In Section 2.4 I will explain how, even if we accept Bivalentism and classical logic, we are still not committed to Determinism. Although Bivalentism is inconsistent with the supervaluationist and multivalentist explications of indeterminacy, there is, I will argue, a coherent understanding of indeterminacy that is consistent with both Bivalentism and classical logic. This view is neither supervaluationist nor multivalentist, and so escapes Williamson’s critiques of these positions. I will argue in Sections 2.5, 2.6 and 2.7 that this view is a more plausible

Bivalentist view than Determinism. In Section 2.8, I will describe Williamson's Omniscient Speakers Argument. I will explain how, as Mario Gómez-Torrente (1997) has pointed out, this argument is question-begging.

In Section 2.7, in particular, I will address the question of what *indeterminacy* consists in. I will offer an account of indeterminacy, in Subsections 2.7.3-2.7.7, that is non-circular and non-epistemic. In Chapter 3, we shall see how a proper understanding of semantic indeterminacy sheds light on the debate between the so-called Simple and Complex Views about diachronic personal identity. And in Chapter 4, a proper understanding of indeterminacy will assist our investigation of the so-called *Fission Case*.

2.2 SORITES PARADOXES AND THEORIES OF VAGUENESS

2.2.1 Sorites paradoxes

Though the first recorded sorites paradox arguably appears in the Bible (Genesis 18: 23-33)²¹, the first time sorites paradoxes appeared in the form of explicitly philosophical puzzles was in the fourth century B.C., when Eubulides of Miletus posed 'the Bald Man' and 'the Heap'. In those days, it was typical for such logical puzzles to be presented as series of questions. (Williamson 1994: 8)

In modern times, sorites paradoxes have generally been presented as arguments, rather than as series of questions. There are a few variant formats, but a simple way of presenting a sorites argument is as follows. The following is the sorites argument that I will use as my standard example throughout this chapter. I will call it 'the Tadpole Paradox'. It is due to James Cargile (1969). It is a useful example, because in it, many of the extraneous considerations that complicate proper analysis of traditional sorites such as the Bald Man are absent.²² The Tadpole Paradox is based

²¹ Williamson (1994: 276 n. 1) cites Barnes, J. 1982. 'Medicine, experience and logic' in Barnes, J. Brunschwig, J., Burnyeat, M. F. & Schofield, M., eds., *Science and Speculation*, Cambridge, Cambridge University Press, for this example.

²² These extraneous considerations include, for example, variations in sense due to context, as in the case of the predicate 'is tall'. Whom we would call 'tall' in the context of a conversation about jockeys would vary significantly from whom we would call 'tall' in the context of a conversation about basketball players. Yet we do not want to conflate this property of context-sensitivity with that of vagueness, through our use of such an example as 'tall'. In the case of the predicate 'bald', the
(footnote continued next page)

on the supposition that we have before us a tadpole – call him ‘Amphibius’ – which transforms in the usual gradual fashion into an adult frog, over a period including the series of instants T(1), T(2), ..., T(43,545,600), each of these instants occurring exactly one twenty-fourth of a second after its predecessor. (Cargile suggested such intervals in order to capture the notion that we might record the life of this creature on standard film, with each frame corresponding to one of these instants.) At T(1), Amphibius is clearly a tadpole. At T(43,545,600), it is clear that Amphibius is not a tadpole, but a frog. We can construct the following sorites argument based on this *sorites series* of instants.

- | | |
|------------------|--|
| (Premiss 1) | Amphibius is a tadpole at T(1). |
|
 | |
| (Pr.2) | If Amphibius is a tadpole at T(1), then he is a tadpole at T(2). |
|
 | |
| (Pr.3) | If Amphibius is a tadpole at T(2), then he is a tadpole at T(3). |
|
 | |
| ... | |
| (etc.) | |
|
 | |
| ... | |
| (Pr.43,545,600) | If Amphibius is a tadpole at T(43,545,599), then he is a tadpole at T(43,545,600). |
|
 | |
| <i>therefore</i> | |
|
 | |
| (Conclusion) | Amphibius is a tadpole at T(43,545,600). |

The conclusion is inferred from the 43,545,600 premisses by a series of applications of *modus ponens*.

There are two things to be noted about this argument. Firstly, it is a paradox, in that in it an apparently false conclusion is drawn from apparently true premisses by apparently valid reasoning.

distribution, as well as the number of hairs, may be relevant to the predicate's application. In the Tadpole Paradox, there is no need for wordy provisos about such "non-linear" factors.

The premisses do all appear to be true. If one wants to claim otherwise, one is apparently put in the awkward position of having to say just which premiss is the false one. But if one names any particular premiss as the false one, then one is left with no apparent justification for naming that premiss rather than any other. Why, for example, should Premiss 20,000,567 (say) be the singled out? Why not 20,000,568? Or 19,999,997? In asserting the falsity of any particular premiss, one is asserting of a particular pair of instants, only one-twenty-fourth of a second apart, that Amphibius is a tadpole at the first instant but, in the blink of an eye, has ceased to be a tadpole by the second instant. It would seem absurd to posit any particular such pair of instants. It would seem absurd, because Amphibius's transformation is surely *gradual*, rather than sudden. One might *stipulate* that we should use the word 'tadpole' in such a way that after exactly 19,999,997 twenty-fourths of a second, the word ceases to apply to Amphibius. But then one would arguably be talking about a different predicate – one would not be talking about the predicate 'tadpole' as we do in fact use it. As we *do* use it, it is vague.

If we cannot find fault with the premisses, then, might we find fault with the reasoning? But it seems that there is no escape here either. If the argument is invalid, then either *modus ponens* is invalid, or else we cannot validly chain multiple *modus ponens* inferences. But, as Michael Dummett (1975: 103) has pointed out, the validity of *modus ponens* seems constitutive of the very meaning of the word 'if'; and the transitivity of the validity relation seems equally difficult to deny, given that the chaining of valid arguments together seems essential to the very notion of a logical proof.²³

Perhaps one might instead just accept the conclusion as true, holding that Amphibius is a tadpole for the duration of his existence. But then, we could construct such a sorites argument for *any* predicate that is vague. Recall the series of dogs in Section 1.2, each dog being imperceptibly smaller than its predecessor, the first of

²³ The unrestricted validity of *modus ponens* has of course been questioned. See Meyer, Routley & Dunn (1979) on Curry's Paradox. McGee (1985), too, questions certain instances of *modus ponens*. McGee's 1985 has been criticised by Katz (1999). But in any case, McGee & McLaughlin (1995 n. 1) point out that McGee thinks *modus ponens* is still valid in case the conditional premiss does not have conditionals as constituents; and thus even McGee would take *modus ponens* to be valid in an argument like the Tadpole Paradox. However, the subvaluationist approach to vagueness, which Hyde (1997) argues is no less plausible than supervaluationism, does explicitly reject (an unrestricted version of) *modus ponens*, in the context of sorites paradoxes. In any case, *modus ponens* is at least *prima facie* (footnote continued next page)

which is a Great Dane, and the last of which is a Chihuahua. We might construct a sorites argument on the basis of this series, the conclusion being that even the final dog, a Chihuahua, is a large dog. This would be to trivialise the predicate ‘is a large dog’. Alternatively, we could run the argument in reverse, for the predicate ‘is not a large dog’, the first premiss being that the Chihuahua is not a large dog, the conclusion being that the Great Dane is not a large dog. One could argue along similar lines to the conclusion that all buildings are large; that no buildings are large; that all buildings are small; that all animals are old; that no-one is tall. To accept the conclusions of sorites paradoxes would be to take ‘tall’, ‘small’, ‘large’, and thousands of other such apparently useful predicates, to be completely trivial, for each would have to apply either to nothing, or to everything, in its range of potential cases.

So, sorites arguments, like the one above, are paradoxes.

The second thing to note about the above argument is that it is rather reminiscent of the Combined Spectrum we discussed in Section 1.4. Of course *Parfit's* argument for Personal Indeterminism, based on the Combined Spectrum, is not at all a sorites argument. As I argued in Section 1.5, Parfit (1984) does not employ sorites reasoning. However, a sorites argument may be constructed on the *basis* of the Combined Spectrum. Suppose there are m operations in the Combined Spectrum. Call the “leftmost” operation (where nothing is done to P) ‘ $O(1)$ ’. *Etc.*. The first premiss of the sorites argument would be: ‘ P is one and the same individual as the person on the table after operation $O(1)$ ’. We might abbreviate this premiss as ‘ $S(1)$ ’. Then the second premiss would take the form: ‘If $S(1)$ then $S(2)$ ’. The third premiss would be ‘If $S(2)$ then $S(3)$ ’. And so on, the argument consisting of a chain of conditional premisses, just as in the Tadpole Paradox. The conclusion of the argument would be that P is one and the same individual as the person on the table after operation $O(m)$ (in which all cells are destroyed and replaced with dissimilar cells). The conclusion is clearly false. However it would seem implausible to reject any particular premiss.

The task of diagnosing what is wrong with sorites paradoxes serves as a benchmark in the debate about how vagueness ought to be logically and semantically accounted for. A theory of vagueness must say something satisfactory about sorites paradoxes. Theorists about vagueness hope that by providing a suitable account of

plausible, and hence the Tadpole Paradox is still *apparently* valid – which is all I am trying to establish here.

vagueness, they may explain what is wrong with sorites paradoxes. My aim in discussing multivalentism and supervaluationism will principally be to outline the most well-known theoretical options for dealing with vagueness in a way that accommodates (or purports to accommodate) indeterminacy. My main aim will *not* be either to attack or to defend either of these theories; nevertheless I will discuss some of the criticisms to which these theories have typically been subject.

I shall now describe the three theories of vagueness to have been most prominently defended in recent literature: multivalentism, supervaluationism and the epistemic view. I shall explain how each of these theories proposes to deal with the sorites problem. While the first two theories do this in a way that tries to accommodate indeterminacy, the epistemic view rules indeterminacy out altogether.

2.2.2 Multivalentism

Let us begin with multivalentism. This is really a group of theoretical approaches to vagueness which have in common that they postulate more truth-values than two, the other values generally being treated as truth-values proper, rather than as value “gaps”.²⁴

One form of multivalentism holds that we need just three truth-values to deal with vagueness. The third value, besides True and False, is usually called ‘Indeterminate’ or ‘Indefinite’. Thus, the predications of a vague predicate to those objects to which the predicate determinately applies have the value True; those predications where the predicate determinately fails to apply have the value False; and borderline predications, where the predicate neither determinately applies nor determinately fails to apply, have the value “Indefinite”. This third value allegedly affords a diagnosis of what is wrong with sorites paradoxes. In a sorites argument of the form of the Tadpole Paradox, there will be some conditional premiss, lying on the cusp of the borderline zone, such that its antecedent is True, and its consequent is Indefinite. The premiss itself thus takes the value Indefinite. The chain of *modus ponens* reasoning is thereby broken by denying that this premiss is true. Given three-

²⁴ Some theories are not easily classified. For example, Michael Tye’s (1994) theory has more in common with traditional three-valued approaches, than with traditional “gap” theories, since while Tye wants to think of the middle value as a value gap, rather than a truth-value as such, that “gap” features in truth-tables alongside “true” and “false”; *i.e.* the logical connectives operate upon three values. Dorothy Edgington’s (1997) approach, though multivalent, is non-truth-functional, unlike traditional multivalent approaches.

valued logic, one can deny the truth of this premiss without claiming that it is false, and, therefore, without asserting its negation. Thus one avoids having to assert, *e.g.* in the Tadpole Paradox, that there is an n such that Amphibius is a tadpole at $T(n)$, but is not a tadpole at $T(n+1)$.

Modern defenders of the three-valued approach include Michael Tye (1994) (three-valued at least in spirit²⁵) and the later Hartry Field (2003a, 2003b). (The earlier Field, *e.g.* 1994, 2000, preferred a two-valued approach.)

A basic criticism of this approach is that the divide between the True predications of a predicate and the Indefinite ones is just as problematic for three-valued logic as the divide between the True predications and the False ones is for standard two-valued logic. Adopting a more complex logic, it is argued, serves merely to shift, or to mask, the awkwardness of the sorites, rather than to resolve or to come to terms with it. For it seems just as implausible to suppose that the facts determine a sharp boundary between Truth and Indefiniteness as it does to suppose that they determine such a boundary between truth and falsity.

Some would argue that multivalentism yet remains the right approach, but that the number of truth-values required is not two or three, but rather an infinite number. (See *e.g.* Zadeh 1965, Goguen 1969, Machina 1976.) The thought is that the transition from clear tadpoledom (or baldness or *etc.*) to clear non-tadpoledom is a *gradual* one, which ought to be captured by a similarly gradual structure of truth-values, rather than by such crude stepwise stages as are provided by two- or three-valued systems. Such theorists may be called ‘degree-theorists’, because they hold that to each predication there obtains a *degree* of truth, usually represented as a real number between 0 and 1, inclusive. 1 represents perfect truth, 0 perfect falsity, and the numbers in between represent various intermediate “degrees of truth”.

There is some intuitive appeal in this approach; and it may seem to afford a plausible diagnosis of the sorites problem. One might say of the Tadpole Paradox, for example, that the predication of ‘is a tadpole’ to Amphibius has a very slightly lower number as its truth-value at each stage in the series. One might then say that a conditional statement has a value slightly less than 1 (perfect truth) in case its consequent is slightly “less true” than its antecedent. One might then say that, since some or all of the conditional premisses in the Tadpole Paradox are less than perfectly

true, the Tadpole Paradox is unsound. One might say this and still insist that *modus ponens* is valid, in that it generates a perfectly true conclusion just in case its premisses are all perfectly true. The fact that each of the conditional premisses is at least *almost* true explains our inclination to believe these premisses. However, this does not entitle us to believe the conjunction of all the premisses. On this approach, the sorites problem is reminiscent of the Lottery Paradox²⁶. (Edgington's (1997) degree-theoretic, but non-degree-functional approach draws explicitly on this analogy.)

There are some "technical" criticisms to which standard degree-theoretic approaches have been subject. I shall not describe this "technical" critique here (but see Edgington 1997). However there is also a certain more general criticism facing degree-theoretic approaches, which is that they seem, by their very sophistication, to be unsuited to capturing vague semantics. It seems implausible to suppose that a predicate that is vague, like 'heap', would be such as to specify, when applied to one of its borderline cases, an exact numerical truth-value, such as 0.5763. Why not 0.5762? The assignment of such precise numerical degrees seems singularly inappropriate as a means of dealing with vagueness.

One response the degree-theorist may be tempted to make is to claim that degree-theoretic semantics are intended merely as an idealisation of vague semantics, one which, like many economic models of market behaviour, is enlightening despite embodying certain assumptions, such as that there is a function from statements to numeric truth-values. It may be that the facts do not determine either that 'S' has the value 0.5, or that it has the value 0.6. They simply determine that it has a value *around* 0.5 or 0.6. But is not to ignore this imprecision just to ignore vagueness? To shift vagueness to a higher level, and then to assume that it does not exist, is not to explicate or to understand it. Good economic models may assume some things; but they do not assume the problems they are intended to solve to be already solved.

A better reply for the degree-theorist would be this. Rather than saying that 'the (numeric) truth-value of S' is an idealised way of speaking, it would be better, to say that, for some statement S, 'the (numeric) truth-value of S' has an indeterminate

²⁵ See n. 24 regarding Tye's approach.

²⁶ For each combination of numbers that might be selected on a lottery ticket, I am justified in believing that that combination will not win the lottery. However, I am not justified in believing that for every combination of numbers, that combination will not win the lottery.

denotation. Where S is a borderline predication, it is determinately the case that ‘the truth-value of S ’ denotes neither 0 nor 1; but there is no value v between 0 and 1 such that it is determinately the case that ‘the truth-value of S ’ denotes v ; although there may be many values w between 0 and 1 such that that phrase determinately does not denote w . On this view, we still can speak legitimately of *the* truth-value of S , for example when we want to say that $\sim S$ has a truth-value equal to 1 minus *the* truth-value of S . (For any statement P , ‘ $\sim P$ ’ denotes the negation of P .) The definite article is appropriate because S still has only one truth-value; however the facts may not always determine *which* truth-value this is. Of course, this reply involves admitting that indeterminacy persists at the metatheoretical level, *i.e.* the level of statements ascribing truth-values. One might, as a degree-theorist, explicate this metatheoretic indeterminacy by iterating numeric truth-degrees at this level. It might be said, *e.g.*, that the statement ‘ S has the value 0.56’ has the value 0.3. This may seem a clumsy response, but it is the least clumsy option for the degree-theorist who acknowledges metatheoretic indeterminacy. For the alternative is to characterise the indeterminacy of metatheoretic statements other than by saying something about what truth-degrees they have. One would wonder, in that case, why this non-degree-theoretic way of saying that something is indeterminate could not have been used to say of such *first-order* statements as ‘Fido is a large dog’ that *they* are indeterminate. Rather than characterising the indeterminacy of one type of statement one way, and of another type of statement another, it seems less messy for the degree-theorist just to iterate the device of truth-degrees to characterise the indeterminacy of metatheoretic statements as well as of ordinary ones.

A similar iterative response is available to defenders of three-valued multivalentism, when faced with the aforementioned charge that, in positing a zone of Indefinite predications, they posit two sharp boundaries in place of one. They could deal with the transition from Truth to Indefiniteness just as they dealt with the transition from Truth to Falsity. Where S lies on the cusp between True and Indefinite, the statement S is a borderline case of ‘is True’, and also of ‘is Indefinite’. ‘ S is True’ is thus Indefinite, and so is ‘ S is Indefinite’. This is really to acknowledge that the predicate ‘is True’ is itself vague, and admits of borderline cases.

Not all are happy with the iterative strategy, however. An alternative multivalentist way of responding to metatheoretic indeterminacy has been proposed by Michael Tye (1994). Tye’s theory is essentially three-valued, although he

characterises his middle value, ‘indefinite’, as the absence of a truth-value, rather than a third truth-value.²⁷ (I will capitalise ‘indefinite’ henceforth.) Tye holds that a sorites series will be divided, by the vague predicate in question, into three groups: the cases yielding True predications, the cases yielding False predications, and the borderline cases, which yield Indefinite predications. He thus eliminates any sharp transition between truth and falsity, by positing an intermediate category between the two. But the transition from Truth to Indefiniteness, and from Indefiniteness to Falsity, is dealt with differently. Tye assigns the value “Indefinite” to the thesis that there are predications that are neither True, nor False, nor Indefinite. He reasons that to assign “True” to this thesis threatens to multiply the number of categories *ad nauseam*; but to assign “False” is to posit a sharp boundary between Truth and Indefiniteness, and between Indefiniteness and Falsity, by neatly grouping each predication in a sorites series into one of three categories. (1994: 283)

Tye’s approach seems to me unmotivated, however. Assigning “False” to the thesis that there are predications that are neither True, nor False, nor Indefinite does not entail that there is a *sharp* boundary between any pair of these categories, even if it does entail that there is a last predication in the series that is True, and a first predication that is False. I shall return to the defence of this claim in Section 2.5. Presently, let us simply note that metatheoretic indeterminacy is quite a general phenomenon, or apparent phenomenon, which most, if not all, theories of vagueness must confront. For there would seem to be no point in a sorites series that is determined *by the facts* about the objects in that series (facts like the physical characteristics of Amphibius, or the hair patterns of scalps) to be uniquely suited as the location of a sharp boundary between *any* pair of *alethic values* – where by ‘alethic value’ I mean the following.

An *alethic value* is a truth-value, or a truth-value gap (as in supervaluationism, described below), or some other status with respect to the truth-or-otherwise of statements, such that the alethic values constitute a mutually exclusive, mutually exhaustive set of statement categories.

²⁷ See n. 24 regarding Tye’s approach.

Indeterminacy with respect to alethic value has often been called ‘higher-order vagueness’. I will return to this topic in Section 2.8.

Regardless of whether they can effectively meet the challenge of higher-order vagueness, multivalentists do at least attempt to accommodate indeterminacy. For three-valued theories, the indeterminate statements are those taking the middle truth-value; and in degree-theories, the indeterminate statements are those taking a value between but not including 0 and 1. Multivalentism accepts that statements really can be indeterminate; whether it offers the correct analysis of this indeterminacy is another question.

2.2.3 Supervaluationism

If we reject multivalentism, must we therefore insist that even borderline predications are either true or false? Perhaps we need not. There is a theory which maintains, with some plausibility, that while borderline predications are correctly regarded as neither true nor false, there are yet only two truth-values, properly speaking. This approach is supervaluationism²⁸. Its defenders have included Mehlberg (1958: 256-259), Lewis (1970) and Fine (1975). (Van Fraassen, *e.g.* 1969, developed supervaluational semantics, but did not apply these to vagueness.)

Supervaluationism begins by noting that for any vague predicate – for instance, ‘tall’ – we may *stipulate* one or another precise meaning for that predicate. Suppose, for example, that we want to conduct a statistical analysis of a pool of potential recruits into an Australian football team, with a view to assessing how many of these would be suitable for ruck positions, in which tallness is a desirable characteristic. We might, in prefacing our report on this research, state that ‘for the purposes of this study, ‘tall’ refers to anyone whose height exceeds two metres.’ The stipulation that anyone over two metres in height is to be considered “tall” amounts to a “precisification”, or “way of making precise” the ordinarily vague predicate ‘tall’. Thus precisified, ‘tall’ is no longer a vague predicate. If we went on to explicate further the semantics of this particular precise sense of ‘tall’, we would not be explicating the semantics of a vague predicate; we would not then be dealing with

²⁸ I reserve the term ‘supervaluationism’ for those theories which both (a) employ supervaluational semantics, in the sense that they employ the mechanics of admissible precisifications, and (b) identify truth with “supertruth” (see below). McGee and McLaughlin’s (2004) theory satisfies (a) but not (b); Hyde’s (1997) “subvaluationist” theory also satisfies (a) but not (b).

vagueness. But supervenientism would account for the vagueness of 'tall' by appealing not to any particular precisification, but rather to the fact that 'tall' has a *range of acceptable* precisifications. We might stipulate that anyone whose height is at least 175cm is tall; or we might stipulate that anyone over 185cm is tall; *etc.* There is no single correct precisification. Rather, 'tall' is vague because there are multiple acceptable precisifications. That is, there are many ways in which we *might* make 'tall' precise consistently with the normal vague meaning of 'tall' inasmuch as the stipulated precise meaning does not classify as "tall" anything that we should normally regard as definitely not tall, and vice versa. An acceptable precisification of 'tall' will also respect such so-called 'penumbral connections' (Fine 1975: 124) as the thesis that if x is tall, then anyone of greater height than x is also tall.

How does this concept of acceptable precisifications bear on the status of borderline cases? And how do supervenientists propose to apply this concept to resolving sorites paradoxes? I will present here a simplified account of supervenient semantics: for subtleties and refinements, see Fine (1975). (In the following, assume ' x ' is a precise neutral-context singular term precisely denoting x .) A predication ' Fx ' is said to be "supertrue" if and only if x satisfies ' F ' on all acceptable precisifications of the predicate ' F ', and is "superfalse" if and only if x does not satisfy ' F ' on any acceptable precisification of ' F '. Supervenientists hold that truth *simpliciter* is just supertruth, and falsity *simpliciter* is just superfalsity. Where x satisfies ' F ' on some, but not all, acceptable precisifications, ' Fx ' is neither true nor false, but is without a truth-value. This is what is called a 'value gap'.

For example, consider 'tall'. Under no acceptable precisification of 'tall' is it the case that 100cm high people are "tall". A precisification of 'tall' under which 100cm high people are called 'tall' would not be an acceptable precisification, because it would call people 'tall' who would under the ordinary vague meaning be definitely regarded as non-tall. So, under no acceptable precisification of 'tall' would a 100cm person satisfy 'tall'. Therefore, on supervenientism, '100cm high people are tall' is superfalse and therefore false *simpliciter*. Conversely, '250cm high people are tall' is supertrue and true *simpliciter*, since on all acceptable precisifications of 'tall', 250cm high people count as "tall". That these statements should be false and true respectively seems as it should be. Now, consider a borderline predication of 'tall'. Suppose, for example, we take a man 175cm in height. It would be reasonable to call such a man a borderline case of 'tall'. Under some acceptable precisifications

of 'tall', the man would be counted as "tall", but under others, he would not be so counted. Since the man satisfies 'tall' on some, but not all, precisifications of 'tall', the statement 'This man is tall' is without a truth-value, according to supervenualionism.

In Section 1.2.4 we saw that borderline predications of vague predicates are classic candidates for the appellation 'indeterminate'. A supervenualionist would agree with us here. S/he would hold that borderline predications are indeed indeterminate, and that this indeterminacy may be explicated (at least in part) as a value gap. (But see Subsection 2.7.1, where I ask whether this is really any kind of explanation.)

How is supervenualionism supposed to deal with sorites paradoxes? The supervenualionist thinks that, in the series of predications 'Amphibius is a tadpole at T(1)', 'Amphibius is a tadpole at T(2)', *etc.*, there are three phases. First, there is the band of true predications; then, there is the band of truth-value-less predications; and then there is the band of false predications. The range of truth-value-less predications corresponds to the range of acceptable precisifications: for each such predication, there is some acceptable precisification on which it is true, and some acceptable precisification on which it is false. Between each pair of neighbouring predications in the central part of the series, some acceptable precisification draws its boundary between that pair. Hence, if we take one of the central premisses of the Tadpole Paradox, of the form 'If Amphibius is a tadpole at T(k) then he is a tadpole at T($k+1$)', there is some acceptable precisification on which Amphibius is a tadpole at T(k) but not at T($k+1$), *i.e.* on which the given premiss is false. (There will be others on which the premiss is true.) Thus, the premiss is not true *simpliciter*, since it is not the case that it is true on *all* acceptable precisifications. The Tadpole Paradox would thus be diagnosed by supervenualionists as unsound, since not all of its premisses are true. This is how supervenualionism allegedly resolves sorites paradoxes. Note that the supervenualionist is merely saying that some of the premisses are untrue; s/he does not hold that any of the premisses is actually false.

Note that, *e.g.* with regard to the Tadpole Paradox, the supervenualionist is committed to the truth of:

- (2AA) For some $T(n)$ ($1 \leq n \leq 43,545,600$), Amphibius is a tadpole at $T(n)$ and Amphibius is not a tadpole at $T(n+1)$.²⁹

This is somewhat counterintuitive. It is true under supervaluationism because under every acceptable precisification, there is an n that satisfies this formula. But even though 2AA is true, there is no particular n for which the statement

- (2AB_{*n*}) Amphibius is a tadpole at $T(n)$ and Amphibius is not a tadpole at $T(n+1)$

is true, since there is no single n which marks the boundary on *all* acceptable precisifications. By pointing this out, the supervaluationist may hope to make their commitment to 2AA seem less unpalatable.

Like the degree-theoretic approach, supervaluationism has been subject to some “technical” criticisms. There are questions about the extent to which supervaluationism “preserves classical logic”. There are also questions about whether supervaluationism is capable of retaining an acceptably general variant of Tarski’s disquotational principle of truth. I shall not address such questions here. Keefe and Smith (1997: 29-35) offer a useful, concise summary of critical debate about supervaluationism. However I will note here that supervaluationism must, like multivalentism, say something satisfactory about metatheoretic indeterminacy. It seems implausible to suppose there is a sharp boundary between the true predications and the truth-value-less ones. If there is no such sharp boundary, there should be some predications in the sorites series for which it is indeterminate whether they are true, or whether they are truth-value-less: for they are borderline cases of ‘true’ and of ‘truth-value-less’.

Supervaluationists are not unaware of the challenge of higher-order vagueness. See *e.g.* Fine (1975: 140-150). One common response is the iterative one available to the multivalentists. One recognises that ‘is true’, ‘is truth-value-less’ and ‘is false’ are vague, and characterises their vagueness as one characterises that of other vague predicates: by saying that there are borderline predications of ‘is true’ *etc.*, and these predications are truth-value-less. In other words, ‘is supertrue’ has borderline cases.

²⁹ From now on I will omit ‘($1 \leq n \leq 43,545,600$)’ in such contexts.

How can that be? Well, the predicate ‘is an admissible precisification’ is vague. There are some precisifications for which it is indeterminate whether they are admissible. Where predication ‘ Fx ’ is on the cusp between “true” and “truth-value-less”: there are some, narrower admissible precisifications of ‘admissible precisification’ on which x satisfies ‘ F ’ on all “admissible precisifications”, so that ‘ Fx ’ is (super)true; and there are other, more inclusive admissible precisifications of ‘admissible precisification’ on which, among the “admissible precisifications” of ‘ Fx ’ are precisifications on which x does not satisfy ‘ F ’, so that ‘ Fx ’ is truth-value-less. Thus it is indeterminate whether ‘ Fx ’ is true, or truth-value-less.

Is there any good reason to deny the supervaluationist – or anybody – the iterative response to (apparent) metatheoretic indeterminacy? Some may find the iterative response unpalatable because they are convinced by a certain argument that there must be at least one sharp boundary in a sorites series no matter *how* many times we iterate our preferred “indeterminacy operator”. If this argument is right, iteration does not enable one to escape sharp boundaries, and so is unmotivated. The argument has appeared in different forms, including as Williamson’s (1994) “Omniscient Speakers Argument”. In Section 2.8, I will describe that argument, and show that it is question-begging.

My purpose thus far has not been to defend either supervaluationism or multivalentism, but has rather been to describe two theoretical options for accommodating the thesis that vagueness results in indeterminacy. Next I shall describe how the defender of the epistemic view responds to the sorites. They do so *via* their claim of Determinism. Like multivalentists and supervaluationists, I wish to oppose Determinism; but my critique of it will commit me neither to multivalentism nor to supervaluationism. For we shall see later that these are not the only plausible theories to accommodate indeterminacy.

2.2.4 The epistemic treatment of the sorites

How would a defender of the epistemic view of vagueness respond to the sorites? For such a theorist, the facts either determine that S , or determine that not- S , for any borderline predication ‘ S ’; but we are just ignorant of these facts. A component of the epistemic view is the claim of Determinism. It is this claim on which the epistemic response to the sorites is based.

Recall that I defined Determinism simply as the following thesis.

(Determinism) No statement is indeterminate on account of vagueness.

How would Determinism block the Tadpole Paradox? For each statement of the form ‘Amphibius is a tadpole at $T(n)$ ’, the Determinist is committed to the claim that either the facts determine that Amphibius is a tadpole at $T(n)$, or the facts determine that Amphibius is not a tadpole at $T(n)$. Assuming the Determinist wants to affirm Premiss 1 (that Amphibius is a tadpole at $T(1)$), and to deny that Amphibius is still a tadpole at $T(43,545,600)$, s/he blocks the argument by claiming that one of the conditional premisses is false. Assuming the Determinist wants to retain classical logic and two-valued semantics (the ability to do this being a principal motivation for the position), this entails that:

(2AA) For some n , Amphibius is a tadpole at $T(n)$ and Amphibius is not a tadpole at $T(n+1)$.

2AA may seem somewhat implausible. But recall that supervaluationism is also committed to 2AA. Determinism seems *prima facie* even *more* implausible than supervaluationism, however, because unlike supervaluationism, it is committed to:

(2AC) There is an n such that the facts determine that Amphibius is a tadpole at $T(n)$ and the facts determine that Amphibius is not a tadpole at $T(n+1)$.

To be committed to 2AC is to be committed to the claim that there is a pair $\langle T(n), T(n+1) \rangle$ of adjacent points in the series such that the facts determine of *that* pair that *it* marks the boundary between the points at which Amphibius is a tadpole, and those at which he is not a tadpole. Because the boundary has a unique and *determinate* location, it may be described as a *sharp* boundary. 2AC seems stronger than 2AA, because 2AA is consistent with the claim that, although there is a *boundary*, *i.e.* a pair $\langle T(n), T(n+1) \rangle$ of adjacent points in the series such that Amphibius is a tadpole at $T(n)$ but not at $T(n+1)$, there is not a *sharp* boundary, because, since the facts do not determine of any $\langle T(n), T(n+1) \rangle$ that *it* is the boundary, the facts do not determine the boundary’s location: the boundary, rather, has an indeterminate, “floating” or “fuzzy”

location. (I shall return to this idea in Subsection 2.5.1.) The Determinist, in contrast, in holding 2AC, holds that there is a fact of the matter, of which we are merely ignorant, about the exact point at which Amphibius changes, very suddenly, from a tadpole into a non-tadpole. 2AC in particular is what makes Determinism a difficult thesis to accept, at least on an initial consideration. The word ‘tadpole’ (like other vague predicates) just does not seem to have a sufficiently precise meaning, even in the way it is defined in, say, biological dictionaries, to determine such a sharp boundary in the series. Determinists try to ameliorate the counterintuitiveness of their position by giving an account of vagueness in which predicates can be vague and also have sharp boundaries. They have argued that vagueness is really a kind of *ignorance* about the location of such boundaries.

Determinism should not be dismissed on the mere basis of the strangeness of its immediate implications. The fact that Determinism enables such a simple response to sorites paradoxes is itself a reason to test it further. However the great counterintuitiveness of positing a unique transition point with determinate location means that we should not accept Determinism without some very good arguments in its favour, and without having tested the alternatives and found them wanting. Much of the rest of this chapter will be devoted to assessing Williamson’s (1994) case for Determinism. If this case is strong, then as we have seen, the case for Personal Indeterminism is in trouble, since it relied on the claim that certain SDPIs, on account of the vagueness of their constituent expressions, may be indeterminate. But if the epistemic view is right, vagueness does not generate indeterminacy, and so such statements are not indeterminate; rather, there is a fact about whether they hold, of which fact we are simply ignorant.

Williamson’s case for Determinism has a “negative” and a “positive” component. The “negative” component comprises an extensive critique of rival theories of vagueness, especially multivalentism and supervaluationism. The “positive” component consists of Williamson’s direct arguments in favour of Determinism.

Williamson (1994: 96-164) makes several criticisms of supervaluationism and multivalentism. But I shall not be very interested in whether multivalentism or supervaluationism can be defended. The falsity of these theories would not entail Determinism unless Determinism is the only remaining alternative. However, there is at least one further Anti-Determinist account, which is, as I shall argue in Section 2.4

and subsequent sections, is more plausible than Determinism, and which (see Section 2.7) is consistent with an explication of indeterminacy which is both non-circular and non-epistemic.

I will now describe and critique Williamson's (1994) positive arguments for Determinism. In Section 2.3, I will describe Williamson's³⁰ "Argument for Bivalentism", and will outline some of the options for Non-Bivalentists who want to resist this argument. In Sections 2.4, 2.5, 2.6 and 2.7 I will defend the view that an Anti-Determinist Bivalentist approach to vagueness, which still retains classical logic, is coherent, and is more plausible than Determinism. Such an approach has also been defended by Campbell (1974)³¹, Burns (1991), McGee & McLaughlin (1995, 2004), Wright (at least in his 1995) and the earlier Field (1994, 2000). In Section 2.8 I will describe the "Omniscient Speakers Argument", which argues for the "Sharp Boundary Thesis", a thesis that may seem to sit uneasily with the denial of Determinism. I will explain how this argument is question-begging, as has been pointed out by Mario Gómez-Torrente (1997). I will then briefly discuss the option of accepting the Sharp Boundary Thesis, but denying Determinism.

2.3 WILLIAMSON'S ARGUMENT FOR BIVALENTISM

2.3.1

I hereby define 'Bivalentism' as follows.

(Bivalentism) Every statement is either true or false.

Recall that a *statement* is an *unambiguous strongly indicative utterance type*. (See Subsection 1.2.2.)

³⁰ Burgess (1998 n. 22) notes that 'a compressed version of what is essentially the same argument' as Williamson's Argument for Bivalentism appeared at p. 197 of Burgess, J. A. 1980 *Vagueness and the Theory of Meaning*. D. Phil. Thesis: Oxford.

³¹ Williamson (1994: 300) asserts that Campbell (1974) defends the epistemic view. But Campbell states that he prefers that the uncertainty surrounding borderline predications be labelled 'semantic uncertainty' rather than 'epistemic uncertainty', since 'the uncertainty... is not due to any lack of knowledge'. He argues: 'No amount of knowledge of empirical facts about men or of the meaning of 'short man' will remove the uncertainty' surrounding a borderline case of a short man. (1974: 180)

Let us now examine Williamson's argument for Bivalentism. I will base my exposition on the version of this argument in Williamson (1994: 187-189), which is more sophisticated than that in Williamson (1992).

Before proceeding, I should point out that although I am here describing Williamson's argument *as* an argument *for* Bivalentism, Williamson himself does not present it straightforwardly as an argument for Bivalentism. Williamson claims that his argument shows that 'the supposition of a counterexample to bivalence leads to a contradiction'. (1994: 189) He admits that the argument does not by itself show that a general principle of bivalence must be asserted; only that we must not deny the bivalence of any particular statement. (1992: 146; 1994: 192-193) Williamson presents *another* argument against the adoption of a *neutral* attitude towards bivalence in particular cases. (1994: 193-194) But I will not describe that argument here.

While I have defined 'Bivalentism' in terms of statements, Williamson defines his 'principle of bivalence' in terms of utterances. He does this in order to exclude that weak version of "bivalentism" that asserts that all *propositions* are either true or false, where propositions are understood to be extra-linguistic entities, expressed by means of utterances. This version of "bivalentism" would be consistent with a view according to which expressions such as 'Amphibius is a tadpole at T(20,000)' are neither true nor false, because they fail to express a unique proposition. However, we should not wish to call such a theory 'bivalentist' in the context of the debate about the status of borderline predications. Therefore, Williamson proposes that the bivalence principle that is really in question in the vagueness debate is a principle applying to utterances, which are linguistic entities, rather than to propositions. More precisely, Williamson defines his bivalence principle as the principle that every 'utterance' which 'says that something is the case', *i.e.* which is strongly indicative (see Subsection 1.2.2), is either true or false. (Williamson 1994: 187)

I have defined Bivalentism as the thesis that every statement is either true or false, *i.e.* that every unambiguous strongly indicative utterance type is either true or false. Any unambiguous strongly indicative individual utterance is guaranteed, according to my definition of 'statement', to express a unique statement; and vagueness is not ambiguity. My definition of 'statement' is consistent with the view that a statement need not express a unique proposition. Williamson's (1994) objection to discussing a propositional "bivalence principle", and his motivation for adopting an

utterance-based principle, is based on the possibility of one-to-many mappings from strongly indicative utterances to propositions. (See Williamson 1999: 507.) Given that my definition of ‘statement’ guarantees that each unambiguous strongly indicative individual utterance expresses a unique statement, and given that Williamson (1994: 198) rules out the possibility of ambiguous individual utterances, Williamson should be comfortable with my construal of the bivalence principle (the one that is at issue in the vagueness debate), as the claim that I have called ‘Bivalentism’.³² Certainly, in the event that we are persuaded of Williamson’s bivalence principle, we should also be persuaded of Bivalentism.

Williamson states his bivalence principle carefully. His principle is, strictly speaking, a statement-schema. Strictly speaking, Williamson argues that, for any statement *S* generated from the schema WBP (below) by substituting a strongly indicative sentence for ‘*P*’, and a name of an utterance for ‘*u*’, we should *not deny* *S*.

(WBP) If *u* says that *P*, then either *u* is true or *u* is false.

(Williamson 1994: 187)

The antecedent of WBP is intended to capture Williamson’s qualification that the principle applies only to utterances that ‘say something’ – *i.e.*, in my terminology, strongly indicative utterances. Thus, *e.g.*, an utterance of ‘the moon is made of rock’ will satisfy the antecedent of WBP since that utterance will have said that the moon is made of rock.

Williamson recognises that, in cases of *reference failure*, a grammatically assertoric sentence may be neither true nor false. (1994: 187) Consider, for example, ‘This flock of sheep is grazing’, uttered by an English teacher in a classroom (far from any sheep), to demonstrate the use of collective nouns. ‘This flock of sheep’ would not then refer. But Williamson argues that borderline predications differ significantly from cases of reference failure. (1992: 147-149) Borderline predications, unlike cases of reference failure, *say that something is the case*, and so satisfy the antecedent of

³² Williamson’s understanding of what an utterance is is such that, in case someone speaks or writes in a way that is ambiguous, either: (a) there are really multiple distinct utterances being made simultaneously, one for each distinct meaning; or (b) there is one utterance, but which does not say that
(footnote continued next page)

WBP (and of 2C and of 2D, below). In his formulation of WBP, Williamson intends ‘*P*’ to replaceable by ‘a declarative sentence whose inscription says that something is the case’. (1994: 188) An utterance by TW of ‘I am thin’ will thus satisfy WBP, because this utterance says that TW is thin; whereas the teacher’s utterance does not say that anything is the case, because ‘this flock’ fails to refer.

2.3.2 Williamson’s argument against the denial of any instance of WBP

Williamson argues for WBP – or, rather, against the denial of any instance of it – by *reductio ad absurdum*. Suppose, he says, that someone thinks they have an example of an utterance that falsifies WBP. The sort of utterance opponents of bivalence typically have in mind will be an utterance of a borderline predication. Williamson offers ‘TW is thin’ as an example of a borderline predication, ‘TW’ denoting Williamson himself, who, we are told, is a borderline case of ‘thin’. Suppose, then, that someone alleges that an utterance *u* falsifies WBP. Then it must be that (for some ‘*P*’ the instance of the following holds):

(2A) *u* says that *P*.

(2B) Not: either *u* is true or *u* is false.

Now, having established under the *reductio* supposition that the above must hold, Williamson argues for the following principles about truth and falsity. These are principles that Williamson maintains, not principles that he thinks are entailed by the *reductio* supposition. The principles are really schemata: Williamson claims that we should accept any claim generated from the following two schemas by replacing ‘*u*’ with any name of an utterance, and ‘*P*’ with any strongly indicative sentence.

(2C) If *u* says that *P*, then *u* is true if and only if *P*.

(2D) If *u* says that *P*, then *u* is false if and only if not *P*.

something is the case. (1994: 198) I do not object to Williamson’s understanding of what an utterance is.

Williamson's argument for 2C and 2D is simple:

Given that an utterance says that TW is thin, what it takes for it to be true is just for TW to be thin, and what it takes for it to be false is for TW not to be thin. No more and no less is required. To put the condition for truth or falsity any higher or lower would be to misconceive the nature of truth or falsity. (Williamson 1994: 190)

Williamson argues that the *onus* is upon one opposed to 2C or to 2D to show that there is some other 'constraint' upon the notions of truth and falsity, besides 2C and 2D, which is inconsistent with 2C or with 2D, and is 'at least as important to the notion of truth (or falsity)' as 2C and 2D. (1994: 191-192)

Williamson argues that, from 2A, 2C and 2D, it follows that

(2E) u is true if and only if P

and that

(2F) u is false if and only if not P .

From 2B, 2E and 2F, by substitution of equivalents, it follows that

(2G) Not: either P or not P .

Finally, from 2G, by one of De Morgan's rules, we can derive a contradiction:

(2H) Not P and not not P .

By *reductio ad absurdum*, Williamson concludes that the assumption, that there is an utterance that falsifies WBP, must be rejected.

2.3.3 Williamson's argument as an argument for Bivalentism

To get from the rejection of this assumption to the affirmation of Bivalentism (as I have defined it), would seem straightforward. Since asserting any utterance that falsifies WBP would seem to commit us to a contradiction, we seem led to conclude

that no utterance falsifies WBP. So every utterance that ‘says something that something is the case’ is either true or false. Presumably, every statement (and certainly every statement of much interest, including the SDPIs discussed in Chapter 1) could be tokened by an unambiguous individual utterance that ‘says that something is the case’. Therefore, since every such individual utterance both (a) is either true or false and (b) tokens a unique statement, every statement is either true or false. Thus Williamson’s argument against the denial of any instance of WBP may also be extended, as argument for Bivalentism, as I have defined it. (See Section 2.3.1.)

The overall argument for Bivalentism is a forceful one. Its logic seems watertight and its premisses plausible. However, for one who wishes to deny Bivalentism in the context of vagueness, for example for a supervaluationist or a multivalentist, there are ways of trying to resist the argument. I will now outline the main such ways, without defending any of them.

2.3.4 Questioning the disquotational premisses

The first way is to claim that at least one of the “disquotational” premisses, 2C and 2D, is untrue. This route tends to suit “gap theorists”, *i.e.* those who think borderline predications are truth-value-less – for example, supervaluationists. If ‘*S*’ is a borderline predication, which, presumably, says that *S*, then, according to gap theory: ‘*S*’ is neither true nor false, and so ‘‘*S*’ is true’ is false; ‘*S*’ is not false; since ‘‘*S*’ is true’ differs from ‘*S*’ with respect to truth-value (or, more correctly, with respect to “alethic value”, since ‘*S*’ has no truth-value), the biconditional ‘‘*S*’ is true if and only if *S*’ is untrue; but this biconditional is the consequent of the relevant instance 2C# of 2C* (2C* being the analogue of 2C for statements rather than for utterances, the distinction between statements and utterances being irrelevant for current purposes)

(viz. (2C#) If ‘*S*’ says that *S*, then ‘*S*’ is true if and only if *S*);

and so, since the antecedent of 2C# is true, 2C# is untrue.

A trivial variant of the above line of reasoning, for utterances rather than statements, leads to an untrue instance of 2C.

What has just been said explains why gap theory is inconsistent with (the truth of all instances of the scheme) 2C. But as Williamson points out (1994: 190), such an explanation by itself does not constitute an argument that this inconsistency warrants

the rejection of 2C rather than of gap theory. I will not here consider what substantial arguments the gap theorist might offer against 2C. See Simons (1992) and Wright (1995: 135-138) for further discussion of gap-theoretic responses to Williamson's argument (though Wright 1995 is not himself a gap theorist).

2.3.5 Rejecting Bivalentism without denying it

Another possible response to Williamson's argument is to "reject" Bivalentism without asserting its negation, *viz.* by holding that it is indeterminate whether Bivalentism holds. This response has been defended by Burgess (1998), Schiffer (1998) and the later Field (*e.g.* 2003a, 2003b) (but not the earlier Field, *e.g.* 2000). Williamson himself admits that his argument 'does not immediately show that bivalence must be asserted for particular utterances, only that it must not be denied'. (1994: 193) (Thus the move from the claim that there are no falsifying instances of WBP, to the straightforward claim of Bivalentism, is not automatic.) He presents a separate argument (1994: 193-194) against neutrality with respect to whether bivalence holds of particular utterances. I will not describe that argument here, but shortly I will explain why *I* think it implausible to hold that Bivalentism is indeterminate.

Burgess (1998 §4) argues that assuming '*S*' is a borderline predication, it is "indefinite", and so is its negation, and so is the disjunction '*S* or not-*S*'. Thus Excluded Middle ('*S* or not-*S*') is "rejected" in the sense that it is regarded not as true, nor as false, but as "indefinite". Burgess wants to accept the disquotational premisses 2C and 2D. Thus since '*S*' says that *S*, '*S*' is true if and only if *S*, and '*S*' is false if and only if not-*S*. We can thus substitute "'*S*' is true' for '*S*', and "'*S*' is false' for 'not-*S*', *salva veritate*. So, since '*S* or not-*S*' is indefinite, the statement "'*S*' is true or '*S*' is false' is also indefinite, for Burgess, Schiffer *etc.*. By holding that Williamson's bivalence principle (WBP) is indefinite, rather than false, Burgess *etc.* avoid having to deny its consequent, and thus would avoid affirming 2B ('Not: either *u* is true or *u* is false'), which led to the contradiction.

The stability of this response is questionable. To make sense of the claim that '*u* is true or *u* is false' is indefinite *rather than* true or false, one must invoke a third (alethic) value – which one might characterise either as a truth-value (*e.g.* Field

2003a, 2003b), or as a value-gap.³³ But if there is a third *exclusive* value, then Bivalentism would seem *false*, rather than *indefinite*, in which case we would have to assert 2B after all. To affirm the indefiniteness of Bivalentism, one would have to hold that there is a third value (call it 'M'), but such that it is never determinately the case that *both* u is M *and* u is neither True nor False. It would have always to be that, for any statement S, either (a) S determinately is True or (b) S determinately is False or (c) it is indeterminate which value in {True, M} S has, but it has either, or (d) it is indeterminate which value in {False, M} S has, but it has either, or (e) it is indeterminate which value in {True, M, False} S has, but it has one of these. But on this account it is difficult to see what the point of having a third value is at all. If one has to appeal to the notion of indeterminacy regarding *which* alethic value, out of a set of values, a borderline predication has, then why not explicate borderline predications more simply, by using this very same notion to say that a borderline predication S is such that, though (determinately) it is either true or false, it is yet *indeterminate which* truth-value in {true, false} S has? Of course, theorists like the later Field (2003a, 2003b) have their reasons for being uncomfortable both with affirming Bivalentism and with denying Bivalentism. I will address some apparent reasons for being uncomfortable with the former, in Sections 2.4, 2.5, 2.6 and 2.7.

For one seeking to resist the argument for Bivalentism, it seems to me the more promising course is, like the supervaluationists, to reject (instances of) the disquotational schemata 2C and 2D.

2.3.6 Other ways of resisting the argument for Bivalentism

Before I continue, I should mention briefly two other ways of trying to resist, plausibly or otherwise, the argument for Bivalentism. One might reject the inference from 2G to 2H, by rejecting De Morgan's rule. Alternatively, one might bite the bullet, accepting 2H. For the latter response to have any hope of plausibility, it would have to be accompanied by a defence of some *strongly paraconsistent* logic³⁴, in which a statement of the form ' P and not- P ' need not entail everything; however, a

³³ Schiffer (1998, 2000), who thinks that Bivalentism is indeterminate, but countenances no third truth-value, can arguably be described as implicitly appealing to a value-gap.

³⁴ For the distinction between strong and weak paraconsistency, see Hyde (1997: 658), who cites p. 126 of Arruda, A. I. 1989. 'Aspects of the historical development of paraconsistent logic', in Priest, G., Routley, R. and Norman, J., eds. *Paraconsistent Logic: Essays on the Inconsistent*. Munchen: Philosophia. pp. 99-130.

paraconsistent system in which every statement is either true, or false, or *both* true and false, will, in any case, satisfy both Bivalentism (as I have defined it) and WBP.

2.3.7

I wish now to argue that, even if we accept both Bivalentism *and* classical logic – and hence, even if, as well as accepting Bivalentism, we reject any paraconsistent approach to vagueness – we can still coherently deny Determinism, and we still have good reason to do so. This argument will comprise Sections 2.4, 2.5, 2.6 and 2.7. In what follows, I shall use the term ‘Classical Bivalentism’ to refer to the conjunction of Bivalentism with the thesis that no statement is both true and false. By ‘ICB’ (‘Anti-Determinist Classical Bivalentism’ or ‘Indeterminist Classical Bivalentism’, which terms I use synonymously) I shall refer to the following thesis.

(ICB) Determinism is false, and Classical Bivalentism is true.

I will assume that Classical Bivalentism entails the acceptance of classical tautologies and rules of inference, *i.e.* of classical logic.

2.4 THE COHERENCE OF ANTI-DETERMINIST CLASSICAL BIVALENTISM

2.4.1

It may seem that once Bivalentism is established, the epistemic view straightforwardly follows. Here is one line of reasoning that Williamson appeals to, which I will call ‘the Simple Argument’.

Suppose an utterance of ‘TW is thin’ is either true or false. Then since we do not know that TW is thin and we do not know that TW is not thin, we are ignorant of something. Either ‘TW is thin’ expresses an unknown truth, or ‘TW is not thin’ does.
(Williamson 1994: 185)

The further implication is that if ‘S’ is an unknown truth, then there is a *fact* of which we lack knowledge – this fact is the ‘something’ of which we are supposedly

‘ignorant’. And if it is a fact that *S*, then, presumably, the facts determine that *S*. Thus, if the Simple Argument (interpreted as an argument from Bivalentism to Determinism) is correct, Bivalentism entails that, for any ‘*S*’, either determinately *S* or determinately not-*S*; *i.e.* it entails Determinism (and Determinism*).

Presumably Williamson’s phrase ‘unknown truth’ means simply ‘true statement (or utterance or proposition or *etc.*) ‘*S*’ such that we do not know that *S*’. The challenge, then, for one who would deny that Bivalentism implies Determinism, is to explain how a statement ‘*S*’ could be such that:

- (2I) ‘*S*’ is true;
- (2J) we do not know that *S*; and
- (2K) the facts do not determine that *S*, and the facts do not determine that not-*S* (*i.e.* there is no fact of which we are ignorant).

The challenge might be put thus: How can ‘*S*’ be true if there is nothing about the world – no *fact* – that *makes* it true? If truth is not a relation of correspondence between statements, and facts that make them true (“truthmakers”), then what is it?

This is a substantial challenge; although it is a challenge posed by philosophy, not by mere logic. No contradiction is apparent in the claim that ‘*S*’ is both true and indeterminate. The following principle is not *analytically* true.

- (2L) For any statement ‘*S*’, if ‘*S*’ is true then the facts determine that *S*, and if ‘*S*’ is false, the facts determine that it is not the case that *S*.

Nevertheless, 2L might seem a difficult principle to deny. The *onus* is borne by one who would reject 2L, rather than by its defender. And 2L must be rejected by anyone seeking to occupy a position that is neither Determinist nor Non-Bivalentist. Logic permits this ground to be occupied. But for the position to be philosophically plausible as well, the concepts featuring in 2L – especially *truth* and *determination by the facts* – must be explicated in a way that is both plausible, and consistent with the denial of 2L.

2.4.2 Conceptions of truth

Let us begin with truth. Is there a plausible conception of truth according to which the truth of ‘*S*’ does not entail the determinate facthood of ‘*S*’? One affirmative answer to this question is a paraconsistent one (see Hyde 1997), according to which a true statement is indeterminate if it is also false. But I shall not address that answer here. I am interested in whether there is a plausible such conception of truth consistent with *Classical Bivalentism*.

Any conception of truth must ultimately be judged plausible or otherwise by comparing it with principles *about* truth that are intuitively plausible. Here are two intuitively plausible principles about truth. McGee and McLaughlin (1995: 214) call these ‘the correspondence principle’ (“CPT”) and ‘the disquotation principle’ (“DPT”). (Below I define the disquotation principle differently from the way McGee and McLaughlin (1995: 211, 214) define it. I adopt a universally generalised conjunction of a variant of Williamson’s (1994: 188) disquotational principles, *i.e.* of 2C and 2D, above.³⁵)

(CPT) ‘... the truth conditions for a [statement] are established by the thoughts and practices of the speakers of the language, and ... a [statement] is true only if the nonlinguistic facts determine that these conditions are met.’ (McGee & McLaughlin 1995: 214) (McGee and McLaughlin have ‘sentence’ rather than ‘statement’.)

(DPT) For any utterance *u*, if *u* says that *S*, then *u* is true if and only if *S*, and *u* is false if and only if not-*S*.

Consider CPT. The truth conditions for the statement ‘Pope Benedict XVI was baptised on Kangaroo Island’ are the circumstances under which that statement would be true. The truth conditions are established, among other things, by the way people use the phrase ‘Kangaroo Island’, *viz.* to refer to a particular island in South Australia. The truth-conditions seem to suggest that the statement is true if and only if the non-linguistic facts about Benedict XVI, islands, *etc.*, determine that Benedict XVI has had water poured on him in a certain way, *etc.*, on a particular island. In our actual

world, the facts about Benedict's baptism do not determine that these conditions are satisfied; and so, it seems, the statement must be false, for that reason. It seems plausible to suppose that the statement could *only* have been true if the facts *had* determined that these conditions are satisfied. What else could the truth of the statement consist in? CPT, then, seems plausible, *prima facie*.

Now consider DPT. An utterance of 'Pope Benedict XVI was baptised on Kangaroo Island' says that Pope Benedict XVI was baptised on Kangaroo Island. DPT says that (a) that utterance is true if, and only if, Pope Benedict XVI *was* baptised on Kangaroo Island, and (b) that utterance is false if, and only if, Pope Benedict XVI *was not* baptised on Kangaroo Island. DPT seems at least as plausible as CPT.

Both CPT and DPT are *prima facie* plausible. However, if one is an Anti-Determinist and a Bivalentist, one cannot believe CPT. Bivalentism entails that every statement '*S*' is either true or false. Presumably, '*S*' is false if and only if '*not-S*' is true. Thus either '*S*' is true or '*not-S*' is true. Suppose CPT is true. CPT says, in effect, that where '*P*' is true, (a) the truth-conditions of '*P*' are *determined* by the *facts* about language users' thoughts and practices, and (b) the non-linguistic facts *determine* that those conditions are met. Thus, where '*S*' is true, on the supposition that CPT is true, the (linguistic and non-linguistic) facts determine that *S*. Where '*S*' is false, '*not-S*' is true; and thus the facts determine that *not-S*. Therefore, if Bivalentism is true, and CPT is true, then for every statement '*S*', either the facts determine that *S*, or they determine that *not-S*. Thus, the Bivalentist must deny CPT, if s/he is to avoid Determinism.

If we reject CPT, some might argue that we have abandoned any coherent understanding of the concept of truth, since, it might be alleged, correspondence with facts is essential to truth.

This charge may be answered, firstly, by maintaining that we have not abandoned truth as a concept that satisfies DPT. What we have done is to distinguish two concepts of truth: "disquotational truth", being the concept that satisfies DPT; and "correspondence truth", being the concept that satisfies CPT. Field (1994) and McGee & McLaughlin (1995) defend the making of this distinction. As an Anti-Determinist Classical Bivalentist, one identifies truth *simpliciter* with disquotational truth, and

³⁵ 2C and 2D apply to non-English as well as to English utterances; and they accommodate indexicals.

determinate truth with correspondence truth. Where ‘determinately *S*’ is true *simpliciter*, ‘*S*’ is “correspondence true”, *i.e.* determinately true. Analogously, there are also two conceptions of falsity. For ‘*S*’ to be disquotationally false is for its negation to be disquotationally true; and for ‘*S*’ to be determinately false is for its negation to be determinately true.

Two steps are involved in the above strategy. In Step One, it is argued that there are two conceptions of truth. That is, it is argued that the word ‘true’ equivocates (at least, in the absence of sufficient contextual cues) between ‘disquotationally true’ and ‘correspondence true’. In Step Two, it is argued that truth *simpliciter* should be “identified with” disquotational truth. In other words, for the purposes of “speaking strictly”, *i.e.* in that unequivocal way that is conducive to philosophical clarity, we ought to use ‘true’ only when we mean ‘disquotationally true’. What justifies each of these two steps?

2.4.3 Justifying Step One

Step One is not justified merely by the fact that CPT and DPT are not logically equivalent. One might stipulate in a philosophical paper that one will by ‘true’ mean ‘satisfying CPT’, and stipulate nothing more about it. Or one might do the same thing with DPT. But to say that ‘true’ equivocates between these two stipulatable senses is to say that either (a) it is not the case that our ordinary use of ‘true’ distinguishes these senses, or (b) our ordinary use of ‘true’ takes on one sense in some contexts, and another sense in others (and perhaps fails to distinguish in others). That is the claim that needs defending, if one is to justify Step One. And that claim can be defended. For it seems that in some contexts, for example, when I say, ‘Everything Benedict XVI has said, is saying, and will say, is true’, I am *virtually*, for every strongly indicative utterance *u* that Benedict XVI utters, uttering *u*. That is, I am virtually co-uttering with Benedict all of his strongly indicative utterances. I say ‘everything he utters is true’ as a shortcut, instead of uttering all of those utterances myself. And had I uttered them myself, the force of this would have been to say, instead of “‘*S*’ is true”, just ‘*S*’. It is just that I am doing this in a way that does not oblige me to utter non-English sentences, or sentences with indexicals (‘she’ *etc.*) taken out of context; and in a way that enables me to co-utter his future utterances before he even gets around to uttering them. Thus, on this way of using ‘true’, the locution “‘*S*’ is true’ just is a convenient way of saying ‘*S*’. This is using ‘true’ in a disquotational sense. But in

other contexts, we use 'true' in a correspondence sense. Consider that commands are neither true nor false. Thus the command, 'Eat apples!' is not true. In asserting "'Eat apples!' is not true", I am *not* asserting, 'Not eat apples!' I am not here using 'true' in its disquotational sense, then. Arguably, what I am asserting is something like this: 'Eat apples!' is not determined by the facts to be true, in this case because it does not have any truth-conditions or falsity-conditions at all, and so there are no truth-conditions for the facts either to satisfy, or to fail to satisfy.

2.4.4 Justifying Step Two

So much for Step One. What about the Step Two? Why identify truth *simpliciter* with disquotational truth rather than with correspondence truth? A related question is this: If one denies that '*S*' is either determined by the facts to be true, or determined by the facts to be false, then what sense is there in saying that '*S*' is either true or false? This question would be asked of the Anti-Determinist Bivalentist both by the Determinist, and by the Non-Bivalentist. These questions challenge the stability of Anti-Determinist Bivalentism. If this challenge cannot be answered, then one who thinks that Non-Bivalentism is implausible, may think that Determinism is the only plausible remaining option.

Two points may be made in answer to this challenge. The first point is largely a *practical* one. It is a point about what reasons we have for agreeing henceforth to use the word 'true' *simpliciter* to mean, unequivocally, 'disquotationally true'. The point is directed, in particular, at the supervaluationist, who identifies truth *simpliciter* with supertruth rather than with disquotational truth; and at the subvaluationist, who entertains similar semantics, but identifies truth *simpliciter* with truth on at least one acceptable precisification. (See Hyde 1997.) To identify truth *simpliciter* with disquotational truth makes for a simpler, more intuitive characterisation of the connectives. It enables us to characterise 'or', 'and' *etc.* truth-functionally. Thus '*S* or *T*' may be taken to be true just in case either '*S*' is true or '*T*' is true, and false otherwise. The truth of '*S* or not-*S*' can be characterised in this way even where '*S*' is a borderline predication and is hence neither determinately true nor determinately false. (Compare supervaluationism.) Similarly, '*S* and *T*' may be taken to be true just in case both '*S*' is true and '*T*' is true. (Compare subvaluationism: Hyde 1997: 651.)

The second point is this, and is intended, in particular, to explain how there can be sense in saying that '*S*' is either true or false, even though the facts determine

neither that it is true, nor that it is false. The point is that to say just this enables us to explain how two competent speakers may *disagree* over whether a borderline case of a predicate does, or does not, satisfy the predicate – even when both are apparently fully informed of all the relevant facts. As Burns (1991: 47-55) and Wright (1995) point out, to characterise a borderline predication ‘*S*’ as *neither* true nor false gives the impression that the status of ‘*S*’ is settled.³⁶ But borderline predications are most naturally characterised as *unsettled* – as ‘having been left *open*’. (Wright 1995: 139) Competent, rational and fully informed speakers might disagree over *whether* such a predication is true or false. Burns argues that the facts do not seem to make borderline predications anything *other* than true or false. But if two competent speakers could disagree over the truth-value of some borderline predication, this is yet consistent with denying Determinism, if such disagreements can persist between rational persons both of whom know all the facts.

It might be asked in what sense there is disagreement at all, if it is not about the facts. The answer is that even in conditions of full factual knowledge, a disagreement may persist regarding what to *do*, *i.e.* how to *settle* the question of a given borderline case. But however it is settled, the borderline predication is either true, or false. The way in which one settles such a question may have practical implications; and one may take these implications into account in determining how to settle it. The absence of a *fact* of the matter about how one *must* settle it leaves room for the consideration of these practical consequences of settling it in a particular way. This is especially clear in cases of judicial arbitration with respect to such vague legal predicates as, *e.g.* ‘is an earning activity’.³⁷ The disagreement is practical, and its resolution involves deliberation, rather than *merely* the gathering of facts. There may be disagreements over borderline cases in which the gathering of further factual information will assist in settling the disagreement; but that in itself does not show the resolved disagreement to have been wholly factual, non-deliberative and non-pragmatic in nature, since the further facts may have been facts about how some way of settling the question would affect people’s lives, or afford a just outcome, or result in legal rules that are easy to understand and to implement, or *etc.* Furthermore, there are surely some “merely verbal” disputes that no amount of further factual

³⁶ Campbell (1974: 182) makes essentially the same point.

information of *any* kind would help the disputants to resolve. The dispute may then be resolved by one person's giving in or changing his/her mind for the sake of its resolution; or by both parties "agreeing to disagree"; or by a non-rational mechanism such as a fight. In any case, the disagreement is a practical one: it is about how we *should* use words – how we should settle a predicate's extension in a particular case. It is not about how we *do* use words.

So why say that the predication is either true or false? To say, of a predication under this kind of dispute, that it is either true or false, is to characterise it as something we might settle one way or the other. To debate whether 'Jim's five-sheep hobby farm constitutes an earning activity (for tax purposes)' is true or false makes sense (even when it is not sensible), because it might be either. In contrast, to debate whether 'Ouch!' is true or false, is senseless, because it is neither.

In resolving a dispute about a borderline predication, there is a sense in which a *new*, more precise meaning has been assigned to the same old expression. Does this mean that we now have a new predicate? It seems more natural, at least in many cases, to say that the meaning of the *same* predicate has now been settled for this particular case. Suppose we decide to settle a borderline case by agreeing to apply the predicate to it. We can say that the predicate retrospectively applied even *before* we settled it; the facts *then* did not *determine* that it applied; but the facts *now* do. This certainly would be the right thing to say about legal disputes. The courts must settle, for example, whether so-and-so *did* engage in an earning activity in the last financial year. A borderline predication *now* has just *the truth-value it would have if ever we settled it*. If we ever do settle it, it becomes determinate, but its truth-value does not change: it always had the truth-value that it would have if ever we settled it. If we never settle it, then 'the truth-value that it would have if ever we settled it' has an indeterminate denotation: either it denotes truth, or it denotes falsity, but it is indeterminate which.

Where a truth-value is not (or not yet) determined by the facts, room is left for language-users to settle the matter in a way that takes pragmatic, normative considerations into account. This "settling" is part of a dynamic process of give and take between a language and its users, and helps to account for the change of meaning

³⁷ But see Sorensen (2001b) 'Vagueness has no function in law'. I will not address Sorensen's arguments here.

in words over time. Rejecting Bivalentism suggests that a strangely radical semantic shift occurs when a predicate's extension with respect to a particular borderline case is settled; for it implies that in doing so, a sentence that was previously taken *not* to be true, is now taken to *be* true. This suggests a discord between our linguistic practice before the settling, and our practice after it. Of course there has been a change in our practice. But by identifying the truth-value of ' Fx ' with its prospective truth-value in case we ever settle it, we can say that ' Fx ' *always* was true, while also saying that one and the same predicate ' F ' has been used by us both before and after the settling. This enables the settling to be characterised as a refinement in meaning, rather than as a shift. That the settling of a once-borderline predication occurs at a particular time, and not before that time, consists in the fact that the pattern of language use *etc.* that is in place *after* that time, and not before it, is what determines (by that time) (coupled with the extra-linguistic facts) the truth-value of the predication.

Hence, at least some disputes about borderline predications may most naturally be characterised both as *disagreements* (thus supporting Bivalentism) and as being *not wholly over what the facts are* (thus supporting Anti-Determinism).

The ability to characterise disputes over borderline predications as disagreements over truth-value is not a clinching argument in favour of Bivalentism. But it is an argument that can be wielded more comfortably by the Anti-Determinist defender of Bivalentism than by the Determinist defender of Bivalentism. The Determinist would have the facts determining the truth-values of borderline predications at every point in time. The plausible picture of an interplay between language users, and a dynamic language that is capable of local refinements influenced by pragmatic factors, sits uneasily with Determinism. If the Determinist identifies the truth-value of a currently borderline predication with its prospective truth-value in case of refinement, as described above, then s/he would have trouble admitting the possibility that cases may never be settled, since this would involve taking the denotation of 'the prospective truth-value of ' S '' to be indeterminate. But if, on the other hand, the Determinist takes the current truth-value of ' S ' to be determined only by past or present facts, then local settlings of the truth-values of borderline predications would have either (a) to involve our happening upon the previously determined but unknown truth-value of the predication, either unwittingly by fluke, or mysteriously by gaining access to previously unknown and apparently unknowable truth-determining facts, or (b) to effect shifts in the truth-value of the

predication, which does not accord with the plausible picture of such settlements as being local refinements of meaning, rather than shifts in meaning. The Determinist might retort that a further option is open to him/her, *viz.* asserting that the truth-values of unsettled predications which *will* be settled are determined by these future settlements, but that the truth-values of predications never to be settled are determined by past or present facts. Though coherent, this seems a rather *ad hoc* response. It offers very different explanations of truth-value determination for two classes of unsettled predications which, one should think, ought to get the same explanation.

2.4.5

To summarise this Section (2.4), Anti-Determinist Classical Bivalentism (ICB) is coherent, and has some *prima facie* plausibility, because: (a) disquotational truth may be distinguished from correspondence truth (determinate truth); and (b) truth *simpliciter* should be identified with disquotational truth, but not with correspondence truth, because this permits logical connectives to be characterised truth-functionally, and permits a natural characterisation of disputes over borderline predications, and of settlements of such disputes.

McGee & McLaughlin (1995, 2004) are prominent recent defenders of ICB. But Campbell (1974), the earlier Field (1994, 2000), and Wright (at least in his 1995) also defend this approach. Burns (1991) defends in detail a version of this approach, appealing to the idea that speakers effectively equivocate between multiple precise bivalent languages – so that vagueness concerns the pragmatic relation between the precise languages and their users, rather than the semantic relation between the precise languages and their meanings. Even if they do not appeal to it explicitly, it seems that anyone claiming to be an Anti-Determinist Classical Bivalentist must identify truth *simpliciter* with disquotational truth, to retain Classical Bivalentism, and must distinguish disquotational truth from correspondence truth, to avoid Determinism (since otherwise s/he would have to claim that borderline predications are either determined to be true or determined to be false).

2.5 SOME CHALLENGES FOR ANTI-DETERMINIST CLASSICAL BIVALENTISM

2.5.1

I will now consider some challenges to McGee & McLaughlin's (1995) approach raised by Hyde (1995).³⁸ These challenges must be confronted by anyone wanting to defend ICB, given the commitments of the position. I will introduce these challenges by first spelling out the commitments of ICB with regard to certain claims.

First of all, for a sorites series Z with respect to vague predicate ' F ', where ' y^* ' denotes the successor of y in Z , the defender of ICB is committed to the claim that there is an item x in Z such that Fx and not- Fx^* . In the Tadpole Paradox, for example, if we abbreviate 'Amphibius is a tadpole at $T(n)$ ' as ' $A(n)$ ', the defender of ICB must hold that

(2AA) For some n , $A(n)$ and not- $A(n+1)$.

Recall that the supervaluationist and the Determinist are also committed to 2AA. (See Section 2.2.) But both Determinists and defenders of ICB are, unlike supervaluationists, also committed to the claim that

(2AD) For some n , ' $A(n)$ ' is true and 'not- $A(n+1)$ ' is true.

Defenders of ICB try to make 2AD easier to swallow by explaining that 2AD is weaker than the following claim, which they reject.

(2AC) There is an n such that the facts determine that $A(n)$ and the facts determine that not- $A(n+1)$.

Is 2AD weaker than 2AC merely in a formal way? Or is there a real sense in which 2AD is not as implausible as 2AC? There is.

Defending 2AA and 2AD does not commit one to the implausible thesis that the predicate ' A ' has a sharp boundary; but 2AC does commit one to that thesis. 2AA

³⁸ Some of Hyde's concerns are echoed in Schiffer (1999), and also in the later Field (2003a, 2003b).

and 2AD *do* say that there is a *boundary*, *just* in the sense that there is a pair of adjacent points in the series one of which satisfies 'A' and the other of which does not. However, the defender of ICB (and, for that matter, the supervaluationist), in denying 2AC, holds that the facts do not determine, of any particular pair of adjacent points, that *it* is the pair that straddles the boundary. Thus, the facts do not determine the boundary's location: it has an indeterminate location.

This notion of indeterminate location seems quite a natural way of accommodating the intuitive notion of a boundary that is "floating" or "fuzzy", rather than "sharp". Alternative ways of accommodating this intuition are less attractive. To deny that there is any boundary at all, fuzzy *or* sharp, is to let the vague predicate seep away from its paradigm cases unchecked, so that even a 100,000-haired scalp is bald.³⁹ But to try to capture a boundary's fuzziness by appealing to a gradation of numeric truth-degrees, where each predication in the series gets an exact degree, is to paint an implausibly *precise*, rather than fuzzy, picture. For to suppose, *e.g.*, that *the place of onset of the region of degrees that are less than value 0.5* has an exact location in the series, determined by the facts to be *just here*, is surely to draw a *sharp* boundary. To reconcile the assignment of numeric truth-degrees with the notion of a fuzzy boundary, it seems we must still say how this place of onset itself could be left undetermined by the facts. But then the numeric values have done no real work in characterising boundary-fuzziness, and we are left, again, with the notion of indeterminate location.

Thus, a boundary that satisfies 2AA has indeterminate location; and indeterminate location seems a natural way of characterising the fuzziness of a boundary. To be committed to 2AA, then, is not to be committed to sharp boundaries. To be committed to 2AC is, unarguably, to be committed to sharp boundaries. But 2AC is entailed neither by 2AA nor by 2AD, nor by the two jointly. We can see that this entailment fails once we distinguish truth *simpliciter* (disquotational truth) from determinate truth (correspondence truth). Given that (with certain irrelevant qualifications about indexicals *etc.*) "'S' is (disquotationally) true' is equivalent to 'S', we should be no more reluctant to believe 2AD than to believe 2AA. But given that "'S' is (disquotationally) true' is analytically weaker than "'S' is (correspondence) true', we can assert 2AD and deny 2AC.

³⁹ But see Sainsbury's (1990) 'Concepts without boundaries' for a different view.

However – as Hyde (1995: 257-258) points out – 2AA is itself somewhat counterintuitive. After all, the classical contradictory of 2AA is a premiss of the Tadpole Paradox in its shorter form, and if this premiss (2AE) were easy to deny then the argument wouldn't be a *paradox*.

(2AE) For any n , if Amphibius is a tadpole at $T(n)$ then Amphibius is a tadpole at $T(n+1)$.

Almost no-one denies that 2AE is, *prima facie*, tempting. We should expect from anyone who, like the defender of ICB, wishes to deny 2AE, an explanation of why, though false, 2AE looks appealing. Hyde (1995: 258) urges, 'It is exactly this appeal which needs to be addressed by any classical account of vagueness.'

Anti-Determinist Classical Bivalentism has the resources to meet this demand. McGee & McLaughlin (1995: 204-205, 236-242) suggest that part of the explanation of the plausibility of 2AE is that each of its instances should be assigned, by a rational agent, a subjective probability very close to 1. That is, in the longer form of the Tadpole Paradox (considered in Subsection 2.2.1), each premiss of the following form is very probably true – since only one of the very many such premisses is false.

(Pr. k) If Amphibius is a tadpole at $T(k-1)$, then he is a tadpole at $T(k)$.

The plausibility of 2AE arises from the fallacious, though understandable, thought that if it is justified to believe each of ' S_1 ', ' S_2 ' ..., ' S_j ' then it is justified to believe ' S_1 & S_2 & ... & S_j '. A similar fallacy underlies the Lottery Paradox.⁴⁰ In the case of the Tadpole Paradox, the conjunction of all the premisses of the above form Pr. k is classically equivalent to 2AE.

Another part of the explanation for the initial plausibility of 2AE is that it is impossible to *specify* a counterexample to 2AE, *i.e.* any n such that Amphibius is a tadpole at $T(n)$ but not at $T(n+1)$. (McGee & McLaughlin 1995: 207) This is because the facts do not determine of any n that it is such an n . Nevertheless, the facts *do*

⁴⁰ For each combination c_i of numbers that might be selected on a lottery ticket, I am justified in believing the claim ' Fc_i ' that combination c_i will not win the lottery. However, I am not justified in believing the conjunction of every such claim, *i.e.* I am not justified in believing the claim of the form ' Fc_1 & ... & Fc_m ', where c_1, \dots, c_m are all the combinations.

determine that (a) Amphibius is a tadpole at $T(1)$; and they do determine that (b) he is not a tadpole at $T(43,545,600)$. Thus, *via* classical logic, they determine that there *must* be an n such that Amphibius is a tadpole at $T(n)$ but not at $T(n+1)$. The determinate truth of that existential proposition arises not from the determinate truth of any of its instantiations, but rather from the fact that it is entailed logically by other determinately true propositions, *viz.* (a) and (b). In the sense that the determinate truth of an existentially quantified statement does not require a determinately true instantiation, argue McGee & McLaughlin, the existential quantifier is “non-constructive”. (1995: 207) As McGee & McLaughlin suggest, such an understanding of the quantifier is consistent, at very least, with an account of the quantifier in terms of the valid rules of inference in which it features, *i.e.* syntax, rather than in terms of semantics. But it seems possible even still to give a systematic semantic account of the connectives, including the existential quantifier, that is consistent with the requisite “non-constructivity” – for example by borrowing the supervaluationists’ semantics of multiple acceptable precisifications, but identifying truth *simpliciter* with disquotational truth rather than with supertruth (and identifying determinate truth with supertruth). That is, on every acceptable precisification, either ‘ S ’ is (disquotationally) true, or ‘not- S ’ is (disquotationally) true (*i.e.* ‘ S ’ is disquotationally false), simply because, on each acceptable precisification, either S , or not- S . McGee and McLaughlin (1995: 222-236) outline such a semantic account; but they claim that their solution to the sorites does not hinge on the success of it, but succeeds simply in virtue of the plausible separation of disquotational truth from correspondence truth.

The defender of ICB is able, then, to explain the intuitive appeal of 2AE consistently with its falsity, by appealing (i) to the high probability, for each of its instances, that that instance is true, and (ii) to a “non-constructive” explanation of the determinate truth of 2AA (the contradictory of 2AE), which appeals to the entailment of 2AA by determinately true statements which are not instances of 2AA, thus explaining our inability to specify falsifying instances of 2AE.

2.5.2 A further challenge, with two aspects

Hyde (1995) makes the following further challenge of anyone who defends such a statement as 2AA, *viz.* the claim that there is an n such that $A(n)$ and not- $A(n+1)$. The challenge can be summed up as follows. Says Hyde, ‘In the absence of any apparent explanation of the non-constructiveness of the existential claim [*i.e.*, on our example,

2AA], in the absence of any explanation of the supposed barrier to knowledge, it is reasonable to remain sceptical of the existence of such unknowable instances.’ (1995: 259) The ‘unknowable instance’ in this case is the n such that $A(n)$ and not- $A(n+1)$. Hyde’s challenge, which is echoed by Schiffer (1999: 498-501), has two aspects. *Aspect One* is the demand to explain non-constructiveness. *Aspect Two* is the demand ‘to [explain] the supposed barrier to knowledge’. I will address Aspect Two first.

2.5.3 Answering the Ignorance Interpretation of Aspect Two

One way of interpreting Aspect Two (faithfully or otherwise), which I will call ‘the Ignorance Interpretation’, is as a demand to explain why we are *ignorant* regarding which n is such that $A(n)$ and not- $A(n+1)$. Hyde’s challenge construed thus can be answered by the defender of ICB as follows. This answer is available to anyone who affirms 2AA (e.g. a supervaluationist) while denying Determinism.

Knowledge – as does ignorance – concerns primarily *facts*, and only in a derivative sense does it concern *statements*. Facts constitute the world, and knowledge is knowledge about the world. *Knowledge that S* is knowledge concerning the *statement* ‘*S*’ only derivatively, in virtue of the correct representation by ‘*S*’ of some aspect of the world. Thus, if we do not know that *S*, and we do not know that not-*S*, this failure to know constitutes *ignorance* only if there is a fact of which we are ignorant. If the facts do not determine that *S*, and they do not determine that not-*S*, then there is no ignorance regarding whether *S*, because there is no fact of the matter regarding whether *S*. Any unargued insistence that there *must* be a fact for us either to know, or to be ignorant of, begs the question against the view that truth need not be identified with determinate truth. Since the defender of ICB alleges no fact of the matter, s/he neither makes, nor is committed to, the claim that we are ignorant about which n is such that $A(n)$ and not- $A(n+1)$. (In contrast, the Determinist *does* have to explain alleged ignorance about borderline predications and boundary locations, because s/he claims that there is a fact of the matter about them.) The defender of ICB does not have to explain any ‘supposed barrier to knowledge’, because s/he denies that there is any fact (about the boundary’s location) epistemic access to which is denied us. My failure to catch a fish in my garden is explained by the absence of fish in my garden. It seems incorrect, or at least odd and misleading, to say that a ‘barrier’ prevents my catching fish there. A barrier suggests something on the other side, to which I am being denied access.

One might then ask: ‘So what is *the truth-value*? Does it exist?’ On a disquotational account, the truth-value is an abstraction from a syntactic device. One should not conclude that it corresponds to ‘a truth’, in any ontic sense. That is just to say: disquotational truth is not correspondence truth.

2.5.4 Answering Williamson’s Simple Argument

In this way, the defender of ICB may reply to the Ignorance Interpretation of Aspect Two of Hyde’s challenge. I will shortly consider another interpretation of Aspect Two, but first I will explain how the reply to the Ignorance Interpretation serves also as a reply to Williamson’s “Simple Argument” from Bivalentism to Determinism, which I quoted earlier, and reproduce below, with labels for its constituent claims.

Suppose [2O] an utterance of ‘TW is thin’ is either true or false. Then since [2P] we do not know that TW is thin and do not know that TW is not thin, [2Q] we are ignorant of something. [2R] Either ‘TW is thin’ expresses an unknown truth, or ‘TW is not thin’ does. (Williamson 1994: 185)

The defender of ICB can reply to the Simple Argument by saying that 2O, 2P and 2R are consistent with Anti-Determinism, but they do not imply 2Q, on the plausible understanding of ‘ignorance’ as ‘there being a relevant fact that we do not know’. And 2R is compatible with the absence of ignorance, providing ‘unknown truth’ *just* means ‘statement ‘S’, which is (disquotationally true) true, such that it is not known that S’.

2.5.5 The Unknowability Interpretation of Aspect Two

The other interpretation, which I will call ‘the Unknowability Interpretation’, of Aspect Two of Hyde’s challenge is the demand to explain *unknowability*. The defender of ICB does not allege, and need not explain, ignorance; but s/he does allege unknowability, in the following sense. S/he alleges that (a) there is an n such that (it is true that) $A(n)$ and (it is true that) not- $A(n+1)$, but also that (b) we cannot know, for any n , that $A(n)$ and not- $A(n+1)$. To explain (a), s/he appeals (to disquotational truth and) to the logical entailment of 2AA by other determinately true statements; and to explain (b) s/he argues that there is no fact of the matter (and that this is consistent with bivalence understood disquotationally). That is to say, the facts that there are do

not determine that the predicate is satisfied, or that it is not satisfied. To point this out is to explain that, since ignorance (*i.e.* lack of epistemic access to the *facts*) is not alleged, it need not be explained. But as an explanation of the *unknowability* of (non-sharp) boundary locations, this seems hollow; for we have yet to be given a satisfactory account of what it *is* for ‘*S*’ to be such that “the facts do not determine that *S*, and the facts do not determine that not-*S*”. The explanation of unknowability, in the absence of such an account, invites the accusation that this concept of *indeterminacy/determinacy* is merely an *ad hoc* manoeuvre designed to enable Classical Bivalentism while avoiding having to explain ignorance.⁴¹ For we have not said very much about the indeterminacy concept so far. We have said that facts constitute the world, or the way the world is. We have said that facts are the things that we know, or are ignorant of, since we know or are ignorant about the world. We seem to have said just enough to characterise “indeterminacy” as being such that my knowing that *S* entails that I know facts, and that these facts determine that *S*. I have argued that borderline predications are indeterminate, *i.e.* that they are such that the facts determine neither that they hold nor that they fail to hold. But it then seems too convenient if we “explain” the unknowability of borderline predications and/or boundary locations *merely* by saying that they are indeterminate. What we need, to make this explanation satisfying, is a non-circular explanation of what we mean by ‘being determined by the facts’. The challenge is to show that our expression ‘it is indeterminate whether *S*’ has content amounting to something other than the thesis that we cannot know either that *S* or that not-*S*.⁴²

2.5.6 Aspect One

If we return now to Aspect One of Hyde’s challenge – the demand to explain the non-constructivity of the existential quantifier – it is clearly another way of getting at this same demand to explain the concept of (in)determinacy. To explain non-

⁴¹ Alternatively, one might describe the allegedly *ad hoc* manoeuvre by defenders of ICB as their ‘proliferating conceptions of truth’ (Hyde: 1995). This is just a different way of wording the same charge of *ad hoc*ery, since the correspondence conception of truth is distinguished from the disquotational conception *via* the determinacy operator.

⁴² Dorr (2003: 110 n. 16) rejects the “inflationary” conception of facthood that I have defended, according to which ‘it is a fact that *p*’ (*i.e.* ‘the facts determine that *p*’) is not analytically equivalent to ‘*p*’, on the grounds that adopting this conception means ‘helping ourselves to the crucial notion of determinacy which we were trying to explain’. This seems to be another way of framing the same charge of *ad hoc*ery.

constructivity would be to explain how, given that it is determinate that there is an n such that $A(n)$ and not- $A(n+1)$, it can possibly be that there is no n for which it is determinate that $A(n)$ and not- $A(n+1)$. What does ‘determinate’ mean here? What semantic work does this word do? If the meaning of the determinacy operator is exhausted merely by its being non-constructive in the context of the existential quantifier (and disjunction), then the operator looks like an *ad hoc*, empty syntactical pawn.

2.5.7

Defenders of ICB, then, as is widely acknowledged⁴³, need an explanation of (in)determinacy, and one which is non-epistemic, *i.e.* does not amount merely to (un)knowability. The Determinist, in contrast, holds that all true statements hold determinately; and so s/he does not have to appeal to the notion of (in)determinacy to distinguish determinate truths from indeterminate ones.

I shall return to the matter of explaining indeterminacy in Section 2.7. But first, I will explain how the Determinist bears a burden of his/her own.

2.6 THE DETERMINIST’S EXPLANATORY BURDEN

2.6.1

The Determinist must explain how it is that the vague predicates we use every day, which seem to have fuzzy boundaries, nevertheless have sharply bounded extensions, even though we seem to have no way of discovering where the sharp boundaries are.⁴⁴

Burgess (1998 §0) puts this challenge to Determinism as follows.

⁴³ This has been argued *e.g.* by Williamson (1994: 194-195), and by Schiffer (1999). The need to explain ‘(in)determinately’ has been acknowledged by Field (*e.g.* 1994: 411), who has tried in Field (2000) to develop an account of (in)determinacy based on a modification of Bayesian subjective probability theory, which account he later rejects in (2003a: 10-11) as ‘pretty hopeless’, in favour of a three-valued approach. McGee & McLaughlin (*e.g.* 2004: 124) are aware of the pressure on ICB from the demand to explain what is meant by ‘determinately’.

⁴⁴ Trying to *settle* a boundary’s location in relation to a particular disputed case (*i.e.* settling whether the predicate applies to that case) is not a futile undertaking. But this is an exercise not of discovery, but of meaning-refinement. The Determinist, however, holds that the boundary is *already* sharp, even prior to such any settling. (Pinning the location of the current boundary on future settlements will not help the Determinist, for not all borderline cases will be settled.) (See Subsection 2.4.4.)

We think (reasonably) that we know what kind of fact could contribute to the determination of sharp boundaries for a vague concept, if it had any. Furthermore, although we should readily concede that there are many scientific facts that contribute towards the determination of the extension of that concept which we do not know, we take ourselves to have a grasp of what the totality of those facts would consist in. But the totality of extension-determining facts do not determine sharp boundaries for vague concepts. It is therefore rational to conclude that vague concepts do not have sharp boundaries. (Burgess 1998 §0)

We can imagine conducting a scientific investigation into thinness, its causes and symptoms; and we can imagine conducting extensive surveys about how the word 'thin' is used; but there seems no hope at all that facts discovered by these methods would ever reveal a sharp boundary for 'thin'. But if *any* facts would reveal such a boundary, *such* facts surely would.

Thus, as Wright (1995: 134) and others have pointed out, the Determinist must explain two things: (i) why we seem doomed never to know where the boundary of 'thin' is, in a given sorites series, and (ii) why, despite this, there is still a *fact* of the matter about where the boundary is.

Defenders of ICB, and Determinists, each have their particular explanatory burden. Which *onus* is more awkward to bear, or difficult to discharge? Let us first consider the Determinist's burden.

2.6.2 Part (i) of the Determinist's burden

Let us begin with part (i) of the Determinist's burden. What might the Determinist say about why we seem doomed never to know where the boundary of 'thin' is? While the Anti-Determinist can appeal (at least, as a beginning of an explanation) to the indeterminacy of the boundary's location to explain its unknowability – by explaining that there is no *fact* about the boundary's location for us either to know or to be ignorant of – this response is not available to the Determinist, who holds that there *is* a fact about the boundary's unique location. Because it is so difficult to believe that our failure to know has to do with any ignorance of such *physical* facts as the precise dimensions of a candidate for 'thin', Determinists, *e.g.* Williamson (1994: 205-209), have defended the claim that what we are hopelessly ignorant of are *semantic* facts.

Thus, even though we know all the physical facts, we are still ignorant of whether x is thin because we are ignorant of the exact shape of the semantic function that takes us from possible totalities of physical facts, to truth-values of ‘ x is thin’. For Williamson, there are certain necessary truths of the form ‘if x has dimensions such-and-such, x is thin’, of which we are ignorant.

Williamson (1994: 205-212) argues that this ignorance is consistent with our understanding the word ‘thin’, and knowing how to use it properly. To understand a term and to use it properly, we must have been ‘completely inducted into a practice’ (211), *i.e.* the practice of using it, but this does not require that we have exactly the same dispositions to use the terms as others do, and it does not require that we know the exact shape of the aforementioned semantic function, which is fixed jointly by the dispositions of *all* speakers, and perhaps also by some other things.

The Anti-Determinist will ask, ‘Yes, but *why* are we ignorant of the semantic function’s shape as it applies to borderline cases of ‘thin’? And why can we apparently *never* figure out the location of the alleged sharp boundary of a vague predicate, even if we know everything about everybody’s thoughts and dispositions regarding the predicate?’ Williamson (1994: 226-234) answers these questions as follows. He appeals to the idea that there are independently plausible epistemic *margin for error principles*, the general form of which he puts thus.

(MEP) ‘ A ’ is true in all cases similar to cases in which ‘It is known that A ’ is true. (Williamson 1994: 227)

The relevant kind and degree of similarity ‘depends on the circumstances’. (227) Consider first the following example, which does not involve vagueness. (This is my example; Williamson’s involves judging the number of people in a crowded stadium.) Suppose we are considering a large glass jar which we know to be full of \$1 and \$2 coins. From our vantage point, it is impossible to tell exactly how much money is in the jar. But we do know some things, such as that there is not exactly \$3 in the jar. Now, suppose that (a) there is not exactly \$30, (b) we believe that there is not exactly \$30, and (c) there *is* exactly \$31. Given (c), and given my limited powers of discrimination from this vantage point, my belief that there is not exactly \$30, though true, is unreliable, since there could very well have turned out to be exactly \$30, and I would still have believed that there is not exactly \$30, because my judgment from this

vantage point is, to a degree, inaccurate. For my beliefs to count as knowledge, then, I need to allow a *margin for error*. The size of the margin depends on the conditions and my abilities. If I am quite bad at making this kind of judgment, then *one* applicable margin for error principle may be that, if I know that there are not exactly n , and $|m-n| \leq 20$, then there are not exactly m .

Margin for error principles apply generally, when we have what Williamson calls 'inexact knowledge'. Their plausibility arises from reasons independent of vagueness. But Williamson argues, as follows, that such principles may be applied to explain the ignorance associated with vagueness.

Let Z be a sorites series with respect to the predicate ' F ', the first item in which determinately satisfies ' F ', the last item in which determinately fails to satisfy ' F ', and let ' x^* ' stand for the successor of x in Z . Let ' Fx ' be the predication of ' F ' of x . For any x in Z , x^* and x are virtually indiscriminable with respect to the qualities that, along with the meaning of ' F ', determine the application, or non-application, of ' F '. Suppose the sharp boundary with respect to F in Z lies between x and x^* . Thus, Fx and not- Fx^* . Why can I not know the boundary's location, *i.e.* that Fx and not- Fx^* ? Suppose I know that Fx . Then it must be that Fx^* . For if not- Fx^* , then my belief that Fx would not have amounted to knowledge after all. My belief that Fx would have been correct merely by fluke, since, had the language-use patterns that, for Williamson, fix the boundary's location been just slightly different (with such a difference being too subtle for me to detect), given the similarity of Fx^* to Fx , the boundary might have been to the other side of x , such that not- Fx ; and my belief that Fx would then have been wrong. Thus, on the assumption that I know that Fx , it must be that Fx^* . But not- Fx^* . Therefore I do not know that Fx . Therefore I do not know that Fx and not- Fx^* . In this way Williamson explains why we cannot know the location of the boundary. (1994: 230-234, esp. 233)

Williamson (1996: 40-41), in response to Wright (1995: 149-150), emphasises that error margin principles also explain why we cannot know the truth-values of predications sufficiently close to the borderline, even if they are not right on the borderline. If x is sufficiently close to the (alleged) sharp boundary, and x satisfies ' F ', then there will be a y just on the other side of the boundary from x , sufficiently close to x , such that not- Fy . Thus we cannot know that Fx , since such knowledge, according to a reasonable margin for error principle, would imply that Fy ; but not- Fy . Put differently: Given that not- Fy , for a y similar to x , had the patterns of language-use

that fix the boundary of 'F' been slightly different, x would *not* have satisfied 'F', and thus any belief that Fx would be insufficiently reliable to constitute knowledge. (See also Williamson 1994: 231.)

Thus, it is only to be expected, argues Williamson, that we should be ignorant regarding predications that are close to the sharp boundary in a sorites series, and, in particular, regarding the location of the boundary itself. In this way, Williamson attempts to answer the challenge: 'Why are we ignorant of the semantic function's shape as it applies to borderline cases of 'thin'? And why can we apparently *never* figure out the location of the alleged sharp boundary of a vague predicate, even if we knew everything about everybody's thoughts and dispositions regarding the predicate?'

Let us first consider whether this "error margin account" that Williamson offers succeeds in discharging part (i) of the Determinist's explanatory burden, *i.e.* explaining why we are apparently unable to know where the boundary of 'thin' (and other vague predicates) lies.

Wright (1995) argues that the account does not explain why we can apparently *never* know the location of its sharp boundaries; for a similar account would serve equally well to explain cases where our inability to know some boundary can be rectified with further investigation. Wright (1995: 151) considers a series of canes, each $1/16''$ longer than its predecessor, the first of which is less than $5'6''$ long, the last of which is more than $6'6''$ long. Consider the predicate 'is less than $6'$ long'. If I am judging the lengths of these canes by the unaided eye, without a measuring implement, then Williamson's error margin account will be able to explain why I do not know, of a cane that is slightly less than $6'$ long, that it satisfies this predicate – since it is sufficiently close to a cane that slightly greater than $6'$ long. But I could remedy my ignorance with the help of a ruler. Thus, appealing to margins for error is consistent with my having *some* way to overcome my epistemic limitation, and so it does not explain why, in the case of vagueness, I apparently have *no* way of overcoming my epistemic limitation. (Wright 1995: 151-152)

Williamson (1996: 41) responds to Wright's criticism by pointing out that vague predicates (we might consider, for example, 'is a short cane'), unlike 'is less than $6'$ long' are not precise, even for the Determinist – they do not 'defer[s] to' any precise definitions. According to Williamson (1996: 41), 'the margin for error principles distinctive of vagueness differ from those distinctive of ordinary perceptual

inexactness' in that our ignorance in the case of vague predicates (e.g. 'is a heap') 'is not of how many grains there are but of whether that many grains can constitute a heap'. What distinguishes vague predicates from precise ones, for Williamson, is that we are ignorant about the exact shape of the semantic function associated with 'thin' *etc.*, which takes us from possible sets of physical facts, to truth-values. For Williamson, this ignorance can still be explained by margin for error principles. Williamson (1994: 231) characterises the boundaries of vague predicates as unstable, though sharp. Very small differences in patterns of language use would effect shifts in the boundary of such a predicate. My true belief that x is thin, where x is close to the boundary for 'thin', could very well have been false, had 'thin' had a slightly different pattern of use from the one it actually has. This is why, unlike with the (comparatively) precise predicate 'is less than 6' long', a mere ruler will not help us find the boundary of 'is a short cane'.

2.6.3 Part (ii) of the Determinist's burden

The trouble for Williamson, however, is that even if his error margin account does discharge part (i) of the Determinist's explanatory burden, it does not even touch part (ii). That is, although Williamson has (purportedly) explained our failure to know where the boundary is, he has not explained why, despite this apparently terminal inability to know, we ought to think there is a *fact* about the boundary's location of which we are *ignorant*. We have seen, in refuting Williamson's "Quick Argument", that we cannot infer, from "'S' is true", and from our not knowing that S , that there is a fact of which we are ignorant.

Williamson (1994: 234) argues that our inability to find a sharp boundary does not constitute evidence that it does not exist. But the *onus* surely lies with the Determinist to show that a sharp boundary *does* exist. What we know about a sorites series (classically) entails only that there is a boundary (*i.e.* a counterexample to the tolerance principle). But if, using the notion of indeterminate location, the claim that there is a boundary can be reconciled with the intuitive idea that that boundary is fuzzy rather than sharp, then what is the motivation for believing, *contra* this intuition, that it *is* sharp?

The heaviness of the *onus* borne by the defender of sharp boundaries is especially evident when we consider cases in which we deliberately set out to create borderline predications. Horwich⁴⁵ (1997: 932) suggests that we might introduce into our language a new predicate, ‘glub’, stipulating that ‘integers greater than 20 are glub and integers less than 10 are not’, but making no further stipulations about its meaning. Consider 15. It is very hard to believe that there is a *unique* sharp boundary between the glub integers and the non-glub ones. ‘Glub’ seems similar, in some respects, to vague predicates of natural language, in that we learn its meaning by learning its paradigm cases. Also, many vague predicates in law are like ‘glub’ in another sense, in that their meanings are stipulated in legislation with the apparent *intention* of creating borderline cases. These cases may be settled by the courts as and when they are confronted, thus settling the boundary’s location in relation to a particular case as it arises. But Williamson’s arguments commit him to the view that each such predicate, if coherent, has a unique sharp boundary from the outset, so that it is determined already, for every possible case, whether the predicate applies, or does not apply, to that case.

Williamson (1997) defends what we might call a ‘default solution’ for such questions as whether 15 is glub. Williamson argues:

To determine which [sharply-bounded] property ‘bald’ refers to, the reference-determining factors must determine of each thing x , time t and possible world w whether x at t in w is to have the property, in other words, whether the ordered triple $\langle x, t, w \rangle$ is to belong to the intension of ‘bald’... As a matter of classical logic, either the reference-determining factors do enough to determine $\langle x, t, w \rangle$ to belong to the intension of bald or they do not. If they do not, then by that very fact they determine $\langle x, t, w \rangle$ not to belong to the intension of ‘bald’, for it cannot belong without being determined to do so by the reference-determining factors... Reference can go by default. (Williamson 1997: 509)

⁴⁵ Horwich’s own (1997) account of such predicates as ‘glub’ seems compatible with ICB. My only substantial disagreement with Horwich is with his insistence that there is a ‘fact’ about the boundary’s location. Given his characterisation of borderline predications as ‘indeterminate’, I am unsure in what sense there could nevertheless be a “fact”, unless ‘it is a fact that S ’ is taken to mean simply ‘it is true that S ’. The Determinist can agree with Horwich, providing his ‘fact’ is understood merely in this way; and, as I shall explain, and as McGee & McLaughlin (2004) have argued, the claim (endorsed by *footnote continued next page*)

The Determinist appealing to the default solution might say one of two things about 'glub'. Arguably, the reference-determining factors *determine* only that integers greater than 20 are glub. Thus all integers less than or equal to 20 would, on this Determinist strategy, be taken to be determinately non-glub. Alternatively, the Determinist might say that integers less than 15, because closer to the paradigm non-glub integers, are non-glub, and the integers greater than 15, because closer to the paradigm glub integers, are glub. The default solution is now applied solely to 15. Since consideration of proximity to paradigms does not 'do enough to determine' that 15 is glub, 15 is thereby, according to the default solution, determined not to be glub. On either account, 15 is determinately non-glub.

Horwich (1997: 933) argues that it is very hard to see how Williamson's margin for error principles could help us explain why we do not *know* that 15 is non-glub. How could the reference of 'glub' be "unstable" in the requisite sense, given the seemingly fixed list of stipulations, which stipulations seemingly exhaust, from the outset, its 'reference-determining factors'?

Perhaps, it might be thought, the Determinist can *avoid* having to explain any ignorance regarding whether 15 is glub, by denying that there is ignorance here at all, in some way that is consistent with Determinism. Perhaps, it might be argued, we *can* know that 15 is non-glub, because we can figure out that the 'default principle' is very plausible. To take this line, the Determinist would have either (a) to abandon the claim that vagueness is essentially ignorance, or (b) to deny that 'glub' is a vague predicate. Another apparent way for the Determinist to avoid commitment to alleging ignorance regarding utterances of '15 is glub' is by denying that the scope of Determinism extends to such utterances, either (c) by denying that 'glub' is a legitimate or coherent predicate at all, or (d) by denying that applications of 'glub' to 15 constitute or express genuine statements. Are any of (a), (b), (c) or (d) feasible responses for the Determinist?

Option (a) is rather awkward. Although the claim that vagueness is ignorance is not logically entailed by Determinism (as I have defined it), it nevertheless is part of the broader "epistemic view" generally defended by Determinists, and is crucial to their claim to be able to explain what vagueness is. Options (b) and (c) appear *ad hoc*.

Horwich) that vague predications supervene on (precise) facts is also something the defender of ICB can happily agree with.

What reason could people *other* than Determinists have for denying themselves the ability to stipulate such predicates as ‘glub’, to stipulate them deliberately so as to be vague, and to treat them as meaningful and coherent? As mentioned earlier, legislators seem frequently to avail themselves of this ability.⁴⁶ Such artificial predicates as ‘glub’ seem, rather than being incoherent, to be, in a certain sense, *incomplete*. And they seem to be vague, possessing, albeit in simply modelled form, just that kind of incompleteness or underspecificity that characterises (even if it is not definitive of) the vague predicates of natural language. Also, the meaning of ‘glub’, as seems to be the case with vague predicates of natural languages, is learnt *via* its paradigm cases. Option (d) would appear *ad hoc* too. If ‘1 is glub’ expresses a statement, why should not ‘15 is glub’ do so too? If I utter, ‘The number you are thinking of is glub’, then surely the question of whether I have asserted something to be the case – *i.e.* of whether I have asserted a statement – has the same answer regardless of *which* number you are thinking of. It only fails to express a statement if you do not have any particular number exclusively in mind – and then only on a certain non-Russellian understanding of ‘the’.

The Determinist, then, cannot plausibly evade commitment to the claim that an utterance of ‘15 is glub’ expresses a statement that is either determined by the facts to be true, or determined by the facts to be false, but such that we are ignorant regarding which it is. Williamson’s “default solution” is the best Determinist attempt to explain how this could be so. But as well as the difficulty of explaining, or explaining away, our apparent ignorance regarding whether 15 is glub, a further objection to the “default solution” remains pressing. Merely to *propose* such a default way of resolving such cases is not to establish that the facts determine that such cases *are* resolved this way. We may eventually agree to *settle* the meaning of ‘glub’ in the case of 15; we may even agree that a certain way of settling this meaning is “objectively the best” way of settling it; but this still would not appear to be a *discovery* of whether 15 is glub; and Williamson is committed to claiming that our use of language even *prior* to such a settling determines whether 15 is glub. Furthermore, we might introduce, simultaneously with ‘glub’, another predicate, ‘brub’, and stipulate only that any $n > 20$ is glub and not brub, and any $n < 10$ is brub and not glub,

⁴⁶ But see Sorensen (2001b) ‘Vagueness has no function in law’. I will not address Sorensen’s arguments here.

and that every integer is either glub, or brub, but not both. To apply the default solution for 'brub', we would have to deny it of 'glub', and vice versa. Williamson might suggest that some other default rule would resolve which of these two predicates should be allowed the first default solution. Perhaps 'brub' should get precedence because 'b' comes before 'g' in the alphabet, or because 0 is "basic", and it is brub. But I suspect that we could suggest a further pair or set of predicates that makes *that* rule problematic to apply. It would become increasingly obvious with each elaboration – if it is not very obvious already – that any further default rules proposed by the Determinist are just his/her arbitrary suggestions, and are not the reflection of any "natural default". Alternatively, the Determinist might claim that such sets of predicates cannot coherently be stipulated in this way. That would be a very *ad hoc* response, for the same reason that responding this way to 'glub' on its own was *ad hoc*. What is wrong with introducing such predicates, apart from that they are awkward for Determinists?

Williamson (*e.g.* 1994: 194-195) would argue that if we have good reasons to believe in Bivalentism, and there is no coherent sense in which '*S*' can be true without it being determinate that *S*, then we are *forced* to think that vague predicates have sharp boundaries, because, he would argue, we are committed to:

- (2AC) There is an n such that the relevant facts determine that $A(n)$
and the relevant facts determine that not- $A(n+1)$.

However, if defenders of ICB can discharge *their* explanatory burden, *i.e.* if there is a plausible non-circular explanation of "indeterminacy" according to which borderline predications are both indeterminate, and either true or false, then Determinism would seem entirely unmotivated, given that pretheoretic intuitions strongly suggest that vague predicates draw fuzzy rather than sharp boundaries, and given that ICB, through the notion of indeterminate location, accommodates this idea of fuzziness.

2.7 THE ANTI-DETERMINIST CLASSICAL BIVALENTIST'S EXPLANATORY BURDEN

2.7.1 Combining supervaluational semantics with ICB

One way of trying to explicate indeterminacy compatibly with Classical Bivalentism is by employing supervaluational semantics, but taking truth *simpliciter* to be disquotational truth, rather than supertruth. One can then explicate determinate truth as supertruth. For any '*S*', since '*S* or not-*S*' is supertrue, '*S* is (disquotationally) true or false' is supertrue, *i.e.* determinately true. This approach is suggested by McGee and McLaughlin (1995: 222-236; 2004). However Field (2000: 10-11; 2003b: 459) argues that supervaluational semantics, as an *explanation* of (in)determinacy, is circular. If we say that the determinate truths are the supertruths, *i.e.* the statements satisfied on all admissible precisifications, then we must explain what it is for a precisification to be "admissible". We might say that a precisification is admissible if it is not inadmissible; and we might say that a precisification is inadmissible if some predicate is satisfied-on-that-precisification by some object that clearly does *not* satisfy this predicate. But what sense could 'clearly' have here, other than 'determinately', or 'knowably'? If it has the latter, we have not avoided the epistemic view; but if it has the former, we have not avoided circularity. Field (2000) goes on to develop a Classical Bivalentist account of (in)determinacy based on a theory of degrees of belief. However he later (Field 2003a: 11) rejects that account as 'pretty hopeless', and changes his mind about Bivalentism.

An initial response for the defender of ICB is to suggest, as an earlier Hartry Field (1994: 411-412) has done, that the notion of (in)determinacy is primitive and cannot be reductively defined, but that this does not entail that we can make no coherent sense of it at all. After all, we cannot give a noncircular explication of the *negation* operator. Nevertheless, Field (1994: 411-412) suggests that, as with negation, there are certain things we can say about the rules for using the "determinately" (or "D") operator, and it is by learning such rules that we come to understand it. *E.g.* he suggests that 'D' obeys the rules that the operator 'necessarily' obeys in the modal logic T.

I think, however, that the defender of ICB can do better at explaining indeterminacy than this, and arguably must do so, if s/he is to distinguish "D" from the Determinist's "knowably" operator, which arguably obeys the same logic. If s/he

can explain indeterminacy plausibly in terms of some other more general notion(s) of which it would be unreasonable to demand a reductive explanation, then his/her explanation will not have fallen short. An initial thought is that this other notion might be modal necessity. Perhaps ‘Facts x, y, \dots determine that S ’ can be analysed as ‘Necessarily, if facts x, y, \dots exist/hold, then S ’. However, such an analysis is problematic for the defender of ICB. Williamson (2004: 116-118) poses the following argument.

2.7.2 Indeterminacy and supervenience

Suppose, for the sake of simplicity, that whether someone is bald depends only on the number and configuration of the hairs on his/her head. Williamson takes this to mean that baldness *supervenies* on these characteristics, so that:

- (2S) For all N and C , necessarily, if one is bald and the number and configuration of one’s hairs are N and C respectively, then necessarily if one is bald and the number and configuration of one’s hairs are N and C respectively then one is bald.⁴⁷

Plausibly, non-baldness also supervenes on hair number and configuration, giving us an analogous principle (2S*) with ‘not bald’ in place of ‘bald’. In any case, McGee and McLaughlin (2004: 126) point out that S5 or ‘a weaker modal logic’ would enable us to derive 2S* from 2S. From 2S and 2S*, we can derive that:

- (2T) For every person a , number N and configuration C , either it is necessary that, if a has N hairs in configuration C , a is bald, or it is necessary that, if a has N hairs in configuration C , a is not bald.

Now, on the proposed analysis of determinacy, we have the following principle.

⁴⁷ I employ McGee & McLaughlin’s (2004: 126-127) wording for the claims 2S, 2T, 2U and 2V; their wording is adapted from, and sometimes matches, Williamson’s (1994: 116-118).

- (2U) To say that the number and configuration of a 's hairs determine that p is to say that necessarily if a has that number and configuration of hairs, then p .

Assuming that every person has some number and configuration of hairs, we can derive the following.

- (2V) For any person a , either it is determined that a is bald or it is determined that a is not bald.

An Anti-Determinist cannot accept 2V.

One cannot appeal to the context-sensitivity of 'bald' to escape the argument: it could easily be adjusted to allow for context-sensitivity. It is generally accepted that vagueness is more than just context-sensitivity; and some vague predicates, such as 'is a tadpole', appear not to be context-sensitive at all. Given any particular context, it is difficult to deny that whether one is bald, in that context, depends only on one's number and configuration of hairs. And it is difficult to see how else to explicate this "dependency" of baldness on hair number *etc.* other than by the principle 2S, or something very like it.

McGee and McLaughlin come to the reasonable conclusion that the defender of ICB – or the supervenientist, for that matter – must reject 2U. Thus, s/he must explain (in)determinacy other than in terms of modal necessity. McGee and McLaughlin accept the left-to-right direction of 2U, but reject the right-to-left direction. They claim that supervenience (2S) 'assures us [only] that the number and configuration of a 's hair [*sic*] determines whether a is bald *if anything does*'. (2004: 127) If the number and configuration of my hairs are N and C , and I am bald, then in every possible world in which I have that number and configuration of hairs, I am bald; but that does not entail that that hair number and configuration *determine* that I am bald. Possible worlds differ only in respect of the facts that hold in them, independently of language. In the supervenientist terms endorsed by McGee and McLaughlin, each admissible precisification of our language is such that if, on that precisification, someone who is bald in the actual world has N hairs in configuration C , then that same precisification *would have it or take it that* everyone with N and C , in every possible world, is bald (assuming contextual factors are held constant). The

variation that constitutes vagueness is between the different admissible precisifications of our language, rather than between different worlds. A possible world in which I am bald is always a *different* world from one in which I am not bald. But, if I am a borderline case of baldness, then one and the same world may be described in two different ways: on one admissible precisification of the language, I may be described as ‘bald’; and on another, I may be described as ‘not bald’. 2S just says that *on any admissible precisification*, on any possible world, if one satisfies ‘bald’ and has *N* and *C*, then in any possible world in which one has *N* and *C* one also satisfies ‘bald’ *on that precisification*. This is a so-called ‘penumbral constraint’ (McGee & McLaughlin 2004: 127): a precisification that does not respect such constraints is not admissible.

The distinction between *determining* and *necessitating* (entailing) is an important one. To make it enables one, for example, to defend the view that facts about diachronic personal identity supervene on particular physical and/or psychological facts, while also allowing for indeterminate statements of diachronic personal identity. That this is so will become apparent in Chapter 3.

But to return to our present problem: (in)determinacy cannot be explained in terms of necessity; it still awaits an explanation.

2.7.3 An account of indeterminacy

I suggest, in a similar vein to Dorr (2003), that what is needed is an account of indeterminacy in terms of pragmatically explicable rules for language use. Dorr (2003: 90-91) sketches an account of indeterminacy in terms of linguistic conventions. I shall presently outline an account of indeterminacy which, like Dorr’s is essentially a pragmatic one, but which, unlike Dorr’s, does not hinge on our being able to draw a principled distinction between what counts as a “convention”, and what does not. (See Dorr 2003: 90) My account meshes in an obvious way with the supervaluational semantics favoured by McGee and McLaughlin (1995; 2004).

I do not offer a complete account of indeterminacy. But in describing the basis for an account, I shall mention some of the ways in which it might be made more sophisticated and expanded; and I hope I will have done enough to show that the notion of *indeterminacy* admits of an intuitive, plausible, non-circular and non-epistemic explication.

I introduce my account by explaining first just the notion of an acceptable precisification of a lone predicate. But it should be noted that this is a simplification. A full account of indeterminacy will explain it in terms of acceptable *joint* precisifications of all the expressions in a statement (or, perhaps, of all the expressions in a related string of statements uttered in a single conversation), in order to respect so-called “penumbral connections”. (See Fine 1975: 124-125.) I shall return to this issue a little later.

Assuming that ‘ x ’ is not vague, a predication ‘ Fx ’ is determinately true if and only if x satisfies ‘ F ’ on all acceptable precisifications of ‘ F ’; determinately false if and only if x satisfies ‘ F ’ on *no* acceptable precisification of ‘ F ’; and indeterminate (but either true or false) if and only if it is neither determinately true nor determinately false. (It is either true or false because, on the disquotational conception of truth and falsity defended earlier, ‘ Fx ’ is true or ‘ Fx ’ is false’ is satisfied on all acceptable precisifications, since ‘ Fx or not- Fx ’ is satisfied on all acceptable precisifications.)

An *application rule* $\langle 'F', \varphi \rangle$ of a predicate ‘ F ’ is a rule of the following form, where φ is some precise property, *e.g.* the property of being at least 150cm tall.

$\langle 'F', \varphi \rangle$ For any x : believe ‘ x is F ’ (*i.e.* believe that ‘ x is F ’ is true) if, and only if, you believe that x has φ ; and believe ‘ x is not F ’ (*i.e.* believe that ‘ x is F ’ is false) if, and only if, you believe that x does not have φ .

I suggest that ‘ F ’ can be *acceptably precisified* as ‘has the property φ ’ if and only if it is *not* the case that it is generally assumed⁴⁸ by speakers of a language that speakers of that language *do not* follow the application rule $\langle 'F', \varphi \rangle$. That is not to say that one *must* follow an application rule at all when using ‘ F ’. But one might. Perhaps one might follow more than one such rule, or follow one rule at one time and another at another time. In actuality, it is extremely likely that one has not decided on any particular rule, unless one is being forced to draw a line in a “forced march” sorites series, to avoid asserting an absurdity. (See Section 2.8.) But for any predicate, there are certain application rules which people always assume that you are *not* following.

For example, people assume that you are not following the rule <'is bald', the property of having less than 100,000 hairs>.

I will not fully explain what it is to *follow* an application rule. One way in which my account could be refined would be by explaining the notion of following such a rule in more detail. Perhaps this is better understood as a broader notion, where following a rule does not require consciously following it, or where following some rule for '*F*' does not require following *no other* rules for '*F*'. Such a broad explication would make more plausible the claim that people often follow such application rules. Or perhaps the notion of following an application rule is better understood quite narrowly, requiring the conscious adhering to certain criteria. But consider that, regardless of whether we adopt a broader or narrower understanding of the notion of following an application rule, we can have a reasonable idea about when a person's thoughts and linguistic behaviours are *inconsistent* with the proposition that they are following a given application rule. And what is important for the present account of indeterminacy is that many of the assumptions that people have about how others use, and think about, a given vague predicate can fittingly be described as assumptions constituted by a *ruling out* of the possibility that others are following (or will follow) certain application rules. This will become clearer shortly.

Following an application rule means that one may believe '*x is F*' for some borderline case *x* of '*F*'. One might believe, *e.g.*, '*a is tall*', since (one believes that) *a* is 176cm in height, and one is following the rule <'is tall', the property of being at least 175cm in height>. How, one might ask, could this be consistent with one's recognising, as one surely should, that *a* is a *borderline* case of 'is tall', such that it is irredeemably *unclear* whether *a* is tall? The answer is that, in following the application rule, one adopts a precisification of 'is tall' in one's *idiolect*; but one may do this while at the same time recognising that others' idiolects may involve conflicting precisifications. This recognition constitutes a recognition that '*x is F*' is not *determinately* true, even though it is true in one's idiolect. The common language is an amalgam of many idiolects; and an idiolect may itself be vague, since for any vague predicate, the speaker of the idiolect need not adopt any application rule at all. (See Burns 1991 for a defence of an understanding of vagueness as arising from the

⁴⁸ 'Generally assumed' is to be understood in a broad sense here. It is not required that speakers are theorists of each others' linguistic behaviour. But, plausibly, certain assumptions are, and others are
(footnote continued next page)

multiplicity of precise languages with respect to which our common language is unsettled.) I shall presently explain how communication is possible even between people adopting conflicting application rules with respect to some predicate, or between people one or more of whom adopts no particular application rule. My account borrows elements from Dorr's (2003); and I shall use as an example Dorr's (2003) exemplary vague predicate 'pretty full'.

Suppose there is a glass that is between 65% and 70% full of beer. (Suppose there is no head – that there is virtually no fuzziness regarding where the beer starts and ends.) Such a glass is presumably a borderline case of 'pretty full'. Consider a situation in which Jones asks me, 'Is the glass pretty full?' Jones cannot see the glass, but I can, and I know that it is between 65% and 70% full. Jones is interested in how much beer is in the glass. If I am to communicate effectively, I must (a) not *mislead* Jones, and (b) not *lie*, *i.e.* not assert ' x is F ' unless I believe that ' x is F ' is true.

Suppose I follow the application rule <'pretty full', the property of being between 71% and 85% full>. I thus believe ' x is not pretty full'. Now, if I answer 'Yes' to Jones's question, I will be lying. If I answer 'No', I will not be lying – but will I be misleading Jones? What partly underpins how people use the expression 'pretty full' in particular conversations are the beliefs that people have of the form, 'Person a is *not* following the application rule <'pretty full', ψ >.' As I will explain shortly, these beliefs are mutually reinforcing within a linguistic community, and so are fairly stable over time. Thus, Jones assumes that I am not following the application rule <'pretty full', the property of being between 5% and 10% full>. I know that Jones assumes this. In order to communicate effectively with Jones, I must refrain from using words in a way contrary to his assumptions about how I am using them, and so I must refrain from following this application rule. But I also know that Jones does *not* assume that I am *not* following the rule <'pretty full', the property of being 65%-85% full>. Because, for all Jones assumes, I could be following *that* rule, I ought not to answer 'No' to his question, even given that I am following the rule <'pretty full', the property of being 71%-85% full>. For all he assumes about how I am using 'pretty full', it could be that I am following the rule <'pretty full', the property of being 65%-85% full>, according to which rule I *should* believe that 'the glass is pretty full' is true. Now, Jones knows that I know of certain application rules

not, *implicit* in people's interpretations of and reactions to other speakers.

that Jones assumes that I am not following them, and of others that Jones does not assume that I am not following them. Jones and I also both know that I do not have to answer either 'Yes' or 'No'. I could remain silent, or say 'Sort of', or 'It's roughly two-thirds full'. Given that he knows I could answer in this non-committal way, Jones will assume, if I answer 'No', that *every* rule which he does not assume that I am not following tells me to believe 'The glass is not pretty full'. He will thus tend to think of the glass as containing rather less beer than it actually does, if I answer 'No'. Similarly, Jones will assume, if I answer 'Yes', that *every* rule which he does not assume that I am not following tells me to believe 'The glass is pretty full'. He will thus think of the glass as containing rather more beer than it actually does, if I answer 'Yes'. Thus, if I answer either 'Yes' or 'No' to Jones's question, I will be misleading him, regardless of which application rule (if any) I am following. As Dorr (2003) argues, my reluctance neither to assert nor to deny borderline predications stems not from any ignorance, but from a desire not to mislead.⁴⁹ (See below for a further defence of the idea that there is no ignorance here.) Even if I am following an application rule that tells me to believe that 'The glass is pretty full' is true, I may still have good reason not to assert 'The glass is pretty full', where that glass is a borderline case of 'pretty full', because I have good reason not to mislead Jones. (I might get around this, of course, by explicitly stating what application rule I am following; but that is to stipulate a precise meaning for 'pretty full', rather than to use 'pretty full' as a vague predicate.)

Assumptions regarding which application rules others are not following are both *self-enforcing* (that people commonly assume *A* tends to cause *A* to continue to be true) and *self-reinforcing* (that people commonly assume that *A* tends to cause *A* to continue to be assumed). This gives them a conventional stability. If I follow an application rule which everyone assumes I am not following, such as <'pretty full', the property of being at least 99% full> or <'pretty full', the property of being at least 1% full>, then I will be either unable to assert, or unable to deny, of anything, 'it is pretty full' (unless I make an explicit stipulation), without either misleading others, or uttering sentences that I do not believe. Thus, people tend to steer clear of application rules that everybody assumes they will steer clear of. The common assumptions

⁴⁹ Dorr (2003) gives a sophisticated explanation of what trying not to mislead someone, in this sense, involves. The glass is full/empty to a certain percentage. My aim is to maximise Jones's degree of
(footnote continued next page)

regarding which application rules are not being followed are thus *self-enforcing*. Again, this is not to say that I must follow an application rule at all in using a predicate; but I might follow one of the ones that people do not assume that I do not follow (and I can do so while still having good reason to refrain from asserting or denying borderline predications, because I have good reasons not to mislead people). Further, since I realise that others too have these assumption-“enforced” reasons to continue to steer clear of rules everyone assumes they will steer clear of, I will tend to continue to assume that they will indeed steer clear of these rules. Thus, the assumptions are *self-reinforcing*.

Thus, the question of which precisifications are acceptable and which are not correlates with the question of which application rules might be followed consistently with the goal of communicating effectively using the vague predicate in question. A range of application rules might be followed consistently with this aim. Even where two people are following conflicting application rules, if each has *not ruled out* (by his/her assumptions) the possibility that the other person is using the particular rule that they are using, then successful, non-misleading communication is still possible, because each person will allow for the possibility of a mismatch between their respective application rules by refraining either from asserting or from denying that the vague predicate in question applies to particular borderline cases, *i.e.* cases in which the non-ruled-out application rules do not all agree about whether the predicate applies. These are the indeterminate predications of the predicate.

2.7.4 Some criticisms pre-empted

It might be alleged that I am characterising indeterminacy in an epistemic way, because I might seem to be holding that indeterminacy “arises” in case two people are ignorant about which application rule the other is using. Certainly, ignorance about others’ exact mental states may be a part of the explanation for why people use predicates vaguely, and use vague predicates the way they do. However, mine is not an epistemic explanation of indeterminacy. Note that it is not required for either party to follow *any* application rule. It may be that no single application rule best characterises someone’s use of a predicate, so that there can be no ignorance about “the rule they are using”. But their use of the predicate will be *inconsistent* with *some*

belief in the proposition that it is full/empty to *that* percentage.

application rules (and consistent with some others). The indeterminacy is constituted not by ignorance about which rules others *are* following, but rather by the patterns of assumptions about which application rules others *are not* following. Since language is communal, indeterminacy of predicates of the language is explained in terms of the *shared* assumptions about which application rules nobody follows. These assumptions are, as explained previously, self-enforcing and self-reinforcing. And the assumptions are partially constitutive of the predicate's meaning for the language community. The explanation of indeterminacy ends with these assumptions, not with an appeal to ignorance.

Another allegation might be that my account cannot explain how private, non-verbalised thoughts might contain indeterminacy, since it appeals to assumptions shared by a *community* of speakers. I do not claim that the account can explain vagueness in non-linguistic mental representation – although perhaps it can.⁵⁰ It is intended primarily as an account of vagueness in natural languages. If I am thinking to myself *in a natural language*, and my thought contains a vague predicate '*F*', and this is the same predicate '*F*' as is used in the language I publicly speak, and I have not adopted any particular application rule for '*F*' (and I need not), then it is still the case that '*F*' can be *acceptably precisified* as φ if and only if it is not the case that it is generally assumed by speakers of my language that speakers of that language do not follow the application rule $\langle 'F', \varphi \rangle$. Thus the indeterminacy of certain predications of '*F*', should I simply think such predications to myself, can still be explained *via* the same account. The indeterminacy inherent in the public language, and explained by reference to commonly held assumptions, infects vague words of the public language whenever they are used – even when I am using them to think to myself. If I do not adopt a particular application rule, then I am thinking using the word in its vague public sense. If I adopt a particular application rule, I am thinking using a precise idiolectal understanding of that word.

One might ask: What if I am thinking in my own private language? Couldn't such a language be vague? If I have invented such a language by defining its words in terms of my native natural language's words, then the vagueness of the latter will infect my private language *via* these definitions. One might now ask: What if I have

⁵⁰ Dorr (2003: 110 n. 13) explains how his own account might be extended to explain vagueness in the "language of thought".

invented a private language “from scratch”, perhaps by training myself to say or to write to myself a particular sound or symbol whenever I see a particular sort of object. I do not propose to give an account of how this radically private kind of “language” might be vague. However, it seems, firstly, that a radically private “language” of this sort, which admits of indeterminacy, would be something so dissimilar to an ordinary language, that one should not demand that an account of indeterminacy in ordinary language must, to be acceptable, apply to such a “language” also.⁵¹ Secondly, Determinism seems especially implausible in the context of such radically private “languages” anyhow. Consider that one might “define” every word of such a “language” in just the kind of incomplete way that ‘glub’ and ‘brub’ (above) were defined, training oneself to utter ‘glub’, ‘brub’ *etc.* for certain numerals, but leaving oneself untrained with respect to how one reacts to other numerals.

Although many terms in the above account of indeterminacy are imprecise, I have avoided circularity because I have not explained indeterminacy by appealing to the looseness of any these terms. And I should not be expected to explain vagueness without using any vague words at all. As Hyde (1994: 40) points out, an explanation of ‘vague’ is not circular simply because it contains vague words, just as an explanation of ‘meaningful’ is not circular just because it contains meaningful words. An explanation of ‘vague’ *would* be circular if it used the term ‘vague’, or used unexplained mere substitutes for that term; and an explanation of ‘indeterminate’ would be circular if it used the term ‘indeterminate’ or unexplained mere substitutes for it. My explanation of ‘indeterminate’ is not circular, even though it uses predicates, such as ‘is generally assumed’, some predications of which may be indeterminate.

2.7.5 Some possible refinements

I noted earlier that the account of indeterminacy simply in terms of acceptable precisifications of lone predicates is a simplification. Strictly, a statement is

⁵¹ In saying this, I am not relying on Wittgenstein’s (1953 §§244-271) “private language argument”. Wittgenstein critiques the idea that there can be a kind of language which is apparently more radically private than the kind I am discussing, namely one that names or describes private sensations inaccessible to others. I am not claiming here either that the Wittgensteinian radically private language, or that the kind of radically private language I have just described, is not possible at all, and my argument does not rely on the proposition that such “languages” do not deserve to be called ‘languages’.

determinately true if and only if it is true on all acceptable *joint* precisifications of the all the expressions it contains. (I expect that the above account could reasonably be extended to cover singular terms, nouns, *etc.*.) For people assume, for example, that if you follow the rule $\langle \text{'far'}, \chi \rangle$ then there are certain rules that you will not follow with respect to 'near'. Thus, one must consider the totality of one's statement, in considering its acceptable precisifications, and this is well-established among proponents of supervaluational semantics. The above account explains when 'has the property φ ' is an acceptable precisification of ' F '. But one way in which the account could be refined, without too much trouble, would be to take into account this holistic aspect of what constitutes an acceptable precisification of an entire statement. In essence, a precisification of a statement will be acceptable if and only if people do not assume that you are not following a rule, or set of rules, that tells you to believe that statement if and only if you believe that precisification of the statement.

Another respect in which this account of indeterminacy could be refined would be explicitly to allow for the context-sensitivity of utterances. A string of words uttered in one context may express a determinately true statement, but the same string may express an indeterminate statement when uttered in another context. (Recall that *statements* are individuated by the context-type in which they are uttered, where contexts differ relevantly in type if they effect different meanings for a single linguistic expression: see Subsection 1.2.2.) Context-sensitivity could be allowed for explicitly without too much trouble. We can say that 'has the property φ ' is an acceptable precisification of (the context-abstracted predicate) ' F ' in context c if and only if it is not the case that it is generally assumed by speakers of a language that, in such a context as c , speakers of that language do not follow the application rule $\langle \text{'F'}, \varphi \rangle$. ' Fx ' is determinately true in c if and only if x satisfies ' F ' on all precisifications that are acceptable in c .

2.7.6

Given the above account of indeterminacy, there is no good reason to hold that a vague predicate ' F ' draws sharp boundaries. The intuition that it does not do so is accommodated by the thesis that its boundary has an indeterminate location. 'There is a boundary for F ' is satisfied on all acceptable precisifications; but there is no n for which ' n is the boundary for F ' is satisfied on all acceptable precisifications. To say this is just to sum up some facts about how people do use ' F '. They use it in

accordance with certain assumptions, which make communication possible without requiring that '*F*' be associated with a unique application rule, or even that it be associated with a unique such rule in each context for each utterer. Communication is thus possible using fuzzily-bounded predicates.

To settle a dispute about a borderline predication's truth-value is to agree to make certain assumptions – consistent with the assumptions already generally made – about how the predicate will from then on be used, which assumptions would render the borderline predication henceforth non-indeterminate. This involves agreeing to rule out all of the application rules (that are not already ruled out), except for those rules according to which one should believe that the predication being settled has the agreed-upon truth-value.

The Determinist holds that there is some sense in which vague predicates draw sharp boundaries. To explain how this can be so, s/he must explain in what sense a path is traced from *each* occasion of a predicate's use to some precise property. ICB does not require that such a path be traced. This is consistent with the supervenience of, *e.g.*, baldness on hair number and configuration. To use a predicate I need not be following any unique application rule. (But I might be.) In what sense could I refer by 'bald' to a precise property, if I am not following a particular application rule? That sense cannot be given by modal necessity, which concerns variation among worlds. Vagueness concerns variation among the manifold sets of rules that might be followed in using symbols to describe those worlds, without violating the assumptions in fact made by others regarding which such rule sets one is not following.

How can it be that I do not know that TW is thin, and I do not know that TW is not thin, and either TW is thin, or TW is not thin, but I am not ignorant of anything? Well, the commonly held assumptions about how 'thin' is used are such that a person who knows all the relevant facts is neither assumed to believe 'TW is thin', nor assumed to believe its negation. For all that is assumed about such a person, s/he might believe *either* – or *neither* – consistently with effective communication, providing s/he bears in mind that asserting borderline predications without qualification tends to mislead people, for reasons explained already. Thus the public, shared sense of the word 'thin' comes with no recommendation about whether to believe 'TW is thin', or to believe its negation, even for one who knows all the facts. It is in this sense that there is no fact about whether TW is thin; and so there is no fact of which to be ignorant.

2.7.7

I want now to return to the topic of “higher-order vagueness”. I remarked in Section 2.2 that the multivalentist and the supervaluationist can avail themselves of the “iterative” response to (apparent) metatheoretic indeterminacy. I suggest that defenders of ICB can avail themselves of this response too. In the terms just outlined, the notion of ‘commonly assumed’ is vague. It is indeterminate just what proportion of a language community has to share a certain assumption, for that assumption to count as ‘commonly assumed’. This vagueness is reflected at the syntactic level by iterations of the indeterminacy operator: it can be indeterminate whether ‘S’ is indeterminate, or whether it is determinately true. Is there any reason to reject this general iterative strategy? I will now consider an argument which might lead some to reject this strategy, or to deny indeterminacy altogether. This is Williamson’s “Omniscient Speakers Argument”.

2.8 THE OMNISCIENT SPEAKERS ARGUMENT AND HIGHER-ORDER VAGUENESS

2.8.1

The conclusion of Williamson’s (1994: 198-201) “Omniscient Speakers Argument” is the claim that any sorites series contains at least one *sharp boundary*, and would do so even if Classical Bivalentism were false. I have previously taken ‘sharp boundary’ to be an intuitive concept, and have argued that ‘boundary with determinate location’, in my sense of ‘determinate’, is a plausible analysis of that concept. I will presently extend that analysis, to explain the notion of a sharp boundary “in a sorites series with respect to a vague predicate”. For the purposes of this explanation I will assume that a boundary is sharp if and only if it has a determinate and unique location. However, the correctness or otherwise of what I say here will not hinge on the correctness or otherwise of the analysis of ‘determinately’ which I proposed in Section 2.7; although it will, admittedly, rely on the assumption that ‘determinacy’ is a meaningful and coherent notion.

In the remainder of this Section, I will use ‘D’ to abbreviate ‘it is determinately the case that’, I will use ‘~’ to abbreviate ‘it is not the case that’ (*i.e.* to

negate), and I will express (rather than name) a predication by juxtaposition of a predicate's name with a singular term or terms.

2.8.2

A sorites series with respect to a predicate is a series of items by reference to which one can construct a sorites paradox with respect to that predicate, such that the predicate clearly applies to the first item in the series, and clearly does not apply to the last item. *E.g.*, the series of dogs described in Subsection 1.2.4 is a sorites series with respect to 'is a large dog'. Where F is a vague predicate and $Z (= \langle x_1, \dots, x_n \rangle)$ is a sorites series with respect to F , one way of claiming that there is a sharp boundary in Z with respect to F is to claim that there is what I shall call a 'sharp boundary for F ' in Z , in other words, a pair $\langle x_j, x_{j+1} \rangle$ of neighbouring items in Z such that DFx_j and $D\sim Fx_{j+1}$. The Determinist is committed to claiming that there is a sharp boundary for F in Z ; and this is one of the things that seems so implausible about Determinism. Now suppose we reject Determinism. Presumably we still hold that DFx_1 and $D\sim Fx_n$. But we hold that there is no sharp boundary for F ; and it is possible to do this either (a) while holding that there is a sharp boundary for ' DF ' in Z , *i.e.* that there is a pair $\langle x_j, x_{j+1} \rangle$ of neighbouring items in Z such that $DDFx_j$ and $D\sim DFx_{j+1}$, or (b) while holding that there is no sharp boundary for ' DF ' in Z , *i.e.* that there is no such pair. Now, on my understanding of the phrase 'sharp boundary with respect to F ', holding that there is a sharp boundary for ' DF ' in Z even while denying that there is a sharp boundary for F in Z is one way of holding that there is a sharp boundary with respect to F in Z . There are other ways of maintaining that there is a sharp boundary with respect to F . In general, a sharp boundary in Z with respect to F is an ordered pair $\langle x_j, x_{j+1} \rangle$ of neighbouring objects in Z such that there is some predicate F^* such that DF^*x_j and $D\sim F^*x_{j+1}$, where F^* is constructed from F using the operator ' D ' (which is synonymous with 'the facts determine (that it is the case) that'), and/or the negation sign, and/or a certain range of other operators pertaining to the "status" of items in the series in relation to F , including, in particular, operators that ascribe truth-values or lack of truth-value, *e.g.* 'it is true that', 'it is false that', 'it is neither true nor false that' and 'it is true to degree 0.35 that'. Thus, one might hold that there is a sharp boundary with respect to F by holding that there is a sharp boundary for ' $DDDF$ '; or by holding that there is a sharp boundary for ' $D\sim DF$ '; or by holding that there is a sharp boundary for ' $\lambda x(\text{it is true to degree } 0.3 \text{ that } Fx)$ '; or by holding that there is a

sharp boundary for ' $\lambda x('Fx'$ is false)'.⁵² (Of course, I do not claim that these ways are mutually exclusive.) These are all ways of holding that there is a sharp boundary with respect to F in Z , according to the way I am now using the phrase 'sharp boundary with respect to F '. I will not define this phrase any more precisely, and I am not obliged to, given that I am here merely trying to make sense of a notion that Williamson (1994: 198-201) and Heller (1996: 177-181) deploy in arguing for a thesis ("SBT", below) which I am inclined to disagree with, this notion being one they do not themselves define nearly precisely. This notion, of a sharp boundary *with respect to* F , is a rather broader, more inclusive notion than the more straightforwardly definable one of a sharp boundary *for* F .

'SBT', or 'the Sharp Boundary Thesis', may be defined as follows.

(SBT) For any vague predicate F , and any sorites series Z with respect to F , there is at least one sharp boundary in Z with respect to F .

2.8.3

Williamson's (1994: 198-201) Omniscient Speakers Argument is best interpreted as an argument for SBT. Williamson does not argue explicitly for SBT. He explicitly argues for the claim that there are 'hidden lines' (201) or 'some sort of ... hidden boundary' (200) in sorites series, *i.e.* hidden to ordinary, non-omniscient speakers, but, it is implied, not hidden to omniscient speakers, who know all the facts. It is clear that SBT or something very like it is what Williamson has in mind. SBT is also effectively what Mark Heller argues *for*, in arguing *against* what he calls 'the Indeterminate Indeterminacy Theory' ('IIT'). (Heller 1996: 177-181) I will not discuss Heller's argument here, suffice to say that it is similar to Williamson's, and is similarly question-begging. Presently I will describe Williamson's argument for SBT – what I call 'the Omniscient Speakers Argument' – before explaining why, as Gómez-Torrente (1997: 238-240) has pointed out, it is question-begging.

⁵² The symbol ' λ ' allows a predicate to be formed from an open sentence. ' $\lambda x(Gx)y$ ' is synonymous with ' Gy '.

2.8.4 The Omniscient Speakers Argument: exposition

I will illustrate Williamson's argument using Cargile's (1969) Tadpole Paradox. (Williamson himself discusses a sorites series with respect to 'heap'.) Recall (Section 2.2) that in this paradox we have a series of instants, and a creature, Amphibius, that gradually transforms, over a duration including these instants, from a tadpole into a frog. We can consider the sorites series here to be the series of instants $T(1)$, ..., $T(43,545,600)$. The vague predicate in question is then 'is an instant at which Amphibius is a tadpole'. Now, we are asked to imagine that we have before us a speaker of English – let's call her 'Omni' – who is omniscient with respect to everything that is relevant to our sorites series, and the application or otherwise of our vague predicate to items in that series. Thus, in our example, we should imagine that Omni is fully aware of all the facts about Amphibius's anatomy, his changes across time, down to the level of his individual cells, and even his atoms. Omni is fully aware of the environment around Amphibius; and moreover, she is fully aware of the semantics of the words 'tadpole', 'instant', 'second' *etc.*, even to the level of knowing all the ways that these words are and have ever been used by anyone that has spoken them, or written them. Anything that could possibly be relevant to questions about Amphibius's tadpoledom, will be among the things about which Omni is all-knowing.

Suppose that having filmed Amphibius's life so that each frame is a picture of the creature at an instant in our series $T(1)$, $T(2)$,... (*etc.*), we show Omni the resulting series of frames, one by one, asking her a series of questions: 'Is Amphibius a tadpole here, at $T(1)$?'; 'Is Amphibius a tadpole here, at $T(2)$?'; *etc.* Omni must answer truthfully, without obfuscation, each of these questions in turn. To the first question she will answer 'Yes', and she will doubtless continue to answer simply 'Yes' for some time. But at some point or other, Omni will say something other than 'Yes' – for otherwise she would have to assert that Amphibius is a tadpole even at $T(43,545,600)$, which she cannot, since he is not then still a tadpole, but an adult frog. So, at some point Omni says something other than 'Yes'; perhaps by saying 'Maybe'; perhaps by saying 'It is now indeterminate whether Omni is tadpole'; or perhaps by saying, 'It is now true to degree 0.99999998 that Amphibius is a tadpole'. But whatever she says, she must eventually say something other than 'Yes'.

There is a certain point in the series at which Omni first says something other than 'Yes'. Now suppose that SBT is false. On this supposition, even though Omni is omniscient and honest, this point does not mark some sharp boundary that is hidden

from ordinary, non-omniscient speakers. The thought is that Omni has an option: there is no unique point at which Omni *must* cease saying 'Yes'. She has some discretion. So, just because she says her last 'Yes' at such-and-such a point, we are not entitled to infer that she *must* have said 'Yes' at this point, on account of an alleged hidden sharp boundary located there.

Now suppose that we have many speakers, all omniscient like Omni. Suppose that we show the frames to all of these speakers, one by one. Suppose that these speakers do, as suggested, have some discretion about where to stop saying 'Yes'. If there is no sharp boundary, then they must have this discretion. Now, we might instruct our panel of omniscient speakers to be as liberal as possible in their use of the predicate 'tadpole'. That is, they must all continue to say 'Yes' for as long as is semantically acceptable according to the meaning of 'tadpole'. In that case, there are two possibilities. The first is that they all stop saying 'Yes' at the same frame. In that case, it seems undeniable that that frame must mark some sort of sharp boundary in the series, of which ordinary speakers are merely ignorant. Alternatively, they do not all stop saying 'Yes' at the same frame. In other words, at least one of the speakers stops saying 'Yes' before the speaker(s) who say(s) 'Yes' for the longest stop(s) saying 'Yes'. In that case the former speaker must not have obeyed the instructions to be maximally liberal, since the latter speaker's(s') actions show that the former could have kept saying 'Yes' for longer. Therefore, it must be that all the omniscient speakers stop saying 'Yes' at the same point; and so there must be a sharp boundary.

Alternatively, we could instruct the speakers to be as conservative as possible, applying 'tadpole' to the minimum number of frames permitted by the semantics of that predicate. We can argue along similar lines as above that all the speakers must, if they are to obey these instructions, stop saying 'Yes' at the same point. Thus, there must at this point be a sharp boundary.

Thus, Williamson argues (in effect), given a sorites series with respect to any vague predicate, if SBT is false, there must be at least two sharp boundaries, apparently hidden to ordinary, non-omniscient users of the predicate, in that series: the boundary between the cases to which that predicate could be said to apply under its most liberal possible application, and the others; and the boundary between the cases to which that predicate could be said to apply under its most conservative possible application, and the others. These boundaries must be such that language users who know all the relevant facts are aware of the exact location of each. But if

this is the case, then the exact location of each boundary must be determined by the facts. But in that case, SBT is true, *contra* the supposition that it is false. By *reductio*, it is concluded that SBT is true.

Two questions invite consideration. First: If SBT is true, does it follow that Determinism is true? Williamson has something to say about this question, and I will address it a little later. But first, I will address another question: Is the Omniscient Speakers Argument a good argument for SBT?

2.8.5 The Omniscient Speakers Argument: critique

Gómez-Torrente (1997: 238-240) has pointed out that the Omniscient Speakers Argument ‘relies on the question-begging assumption that there is a sharp boundary between being *conservative* [in one’s application of the predicate] and not being conservative’ (and between being liberal and not being liberal). Let us see how this is so.

Recall Williamson’s suggestion that we might ask multiple omniscient speakers to be as liberal (or conservative) as possible in applying the predicate. If one of the speakers stops saying ‘Yes’ sooner (or, respectively, later) than the others, we would be entitled, so Williamson argues, to accuse that speaker of not having obeyed our instructions. But are we even entitled to assume that Williamson’s instructions *can* be followed? Take the case in which they are instructed to be maximally liberal. If we are being asked to be as liberal as possible in our application of ‘tadpole’, then we are being asked to make the last point at which we say ‘Yes’ the point m such that it is semantically acceptable to assert ‘ m is an instant at which Amphibius is a tadpole’ but not semantically acceptable to assert ‘ $(m+1)$ is an instant at which Amphibius is a tadpole’. But is there any guarantee that the facts determine where this point is? There is not. And if the facts do not determine this, then an omniscient speaker – one who knows all the *facts* – will still not know, of any point, that it is that point. The Omniscient Speakers Argument asked us to suppose, for *reductio*, that SBT is false, so that the facts do not determine the truth-value of every predication in the series. This meant that for some middle predications, an omniscient speaker, though knowing all the facts, would have discretion as to whether to say ‘Yes’ to the question of whether the predicate applied in these cases, since the facts did not determine the answer one way or another. But given that the application of ‘is a

tadpole' is, according to our *reductio* supposition, not always determined by the facts, is it not equally plausible to suppose that the application of the predicate 'is a point at which it is semantically acceptable to apply the predicate 'tadpole'' is not always determined by the facts? We thus have a picture in which the vagueness of 'is a tadpole' generates a zone of discretion with respect to 'is a tadpole', and the vagueness of 'is an acceptable point at which to apply 'tadpole'' generates a zone of discretion with respect to 'is an acceptable point at which to apply 'tadpole''.

Given the vagueness of 'tadpole', then, an instruction to stop saying 'Yes' at the last point at which Amphibius is a tadpole is such that we might plausibly say one of two things about whether it can be obeyed: (i) it cannot be obeyed, since the expression 'the last point at which Amphibius is a tadpole' assumed there is a definite, *i.e.* determinately located, point in the series which is its unique referent, when the facts determine no such point uniquely; or (ii) it can be obeyed, but the expression 'the last point at which Amphibius is a tadpole' can be given a particular referent only in virtue of the arbitrary decision of the person asked to draw the line – a decision open to discretion, since the facts do not determine a unique correct answer. In either case (i) or case (ii), we cannot infer, from the point at which the speaker does in fact stop saying 'Yes', that *that* point marks a hidden boundary determined by the facts; we can infer only that this is *among* the points at which it was acceptable, *vis-à-vis* the facts and the meaning of 'tadpole', for the speaker to stop saying 'Yes' here.

Thus, Williamson's Omniscient Speakers Argument only works if we have *already* established that the predicate 'is a point at which it is acceptable to apply 'tadpole'' is such that it cannot have indeterminate cases. Thus the argument begs the question against the view that the facts can fail to determine any sharp boundaries at all with respect to a predicate. The argument assumes, rather than establishes, that SBT is true.

Williamson pre-empts a response to the Omniscient Speakers Argument; however the response he pre-empts is not one we need to make in order to defeat his argument. He says: 'Some may feel tempted to repudiate the very possibility of omniscient speakers. To do so is to endorse a very strong form of the view that vagueness is an epistemic phenomenon, for it is to treat ignorance as an essential feature of borderline cases.' (1994: 201) But defeating the Omniscient Speakers Argument does not require denying the possibility of omniscient speakers. An omniscient speaker is one who

knows all the *facts*. As I have argued previously, there is no ignorance involved in my not knowing the truth-value of a borderline predication, because there is no fact of the matter about which truth-value it has.

2.8.6 A refined argument for SBT

It may be tempting to offer the following reply in defence of Williamson. Perhaps the instructions should be understood as asking the speakers to say ‘Yes’ for any case in which there is *any* order of indeterminacy at all with respect to the application of ‘tadpole’. The argument can, indeed, take a certain “refined” form. Williamson (1994) probably presents his argument in terms of omniscient speakers, rather than as below, so as to avoid question-beggingly assuming that classical logic is appropriate in the context of vagueness, given that he would prefer his argument to challenge non-classicists. I have no problem with classical logic; but I still would prefer to deny SBT; therefore I must answer to the following “refined” formulation of the Omniscient Speakers Argument. This is closer to one of the formulations in Heller’s 1996.

(Supposition) SBT is false.

- (Step 1) Either there is a k such that k is the first point (in a sorites series with respect to F) at which there is some order of indeterminacy with respect to F , or there is no such k .
- (Step 2) If there is no such k , then (by classical logic, coupled with the thesis, inherent in the denial of SBT, that there is not a non-gradual transition from cases determinately satisfying F to cases determinately failing to satisfy F), it is determinate that F_n (where n is the last point in the series, to which F clearly does not apply).
- (Step 3) There is a k such that k is the first point at which there is some order of indeterminacy with respect to F .

(Step 4) (Where 'D₃' is short for 'DDD', *etc.*) $k-1$, being the predecessor of k , is such that $D_j F(k-1)$, for any j .

(Step 5) k has a determinate location.

(Conclusion) SBT is true.

2.8.7 Replying to the refined argument

The above argument might look devastating for the defender of ICB, or indeed for the supervenientist, who wishes to deny SBT. But the proponent of this argument must say what 'there is some order of indeterminacy (with respect to F) at m ' means. Presumably, this is equivalent to denying that there is no order of indeterminacy at m . But how should one say that there is *no* order of indeterminacy at m ? The idea expressed in Step 4 is that one can say this by saying that you can append 'D' to ' Fm ' as many times as you like, and still say something true: *i.e.*, for any j , $D_j Fm$. To say that there is some order of indeterminacy at k is, on this understanding, to say that, for some j , not- $D_j Fk$. But to establish Step 5 – that k has a determinate location – one must establish that (2AF) *determinately*, for any j , $D_j F(k-1)$, and that (2AG) *determinately*, for some j , not- $D_j Fk$. Is 2AF entailed by the claim that there is no order of indeterminacy with respect to $k-1$? It is not so entailed on *this* definition of 'there is no order of indeterminacy at x ': for 2AF is not itself of the form 'for any j , $D_j F(k-1)$ '. A similar remark applies to 2AG. The proponent of SBT must, in order to make the argument work, define 'there is no order of indeterminacy at m ' in such a way that 'there is no order of indeterminacy at m ' entails 'it is determinate that there is no order of indeterminacy at m '; but it is hard to see how s/he can do this in a non-question-begging way.

2.8.8 Might an Anti-Determinist accept SBT anyway?

The Omniscient Speakers Argument for SBT is unconvincing. But suppose SBT were true. Would even this give us good reason to be Determinists? Williamson argues that the answer is 'Yes', as follows.

Once hidden lines are admitted, why should a line between truth and falsity not be one of them? After all, Section 7.2 [of Williamson 1994] found the supposition of intermediate cases to be incoherent. Moreover, the failure of non-classical logics to mesh with a satisfactory account of vagueness gives us reason tentatively to return to classical logic. (1994: 201)

Williamson then reminds us of his Argument for Bivalentism. The combination of that argument with the Omniscient Speakers Argument, with Williamson's critiques of particular non-classical theories of vagueness, and with his suggested explanation of ignorance by appealing to error margin principles, is supposed to amount to a convincing case for Determinism. But we have seen that even Bivalentists, embracing classical logic, have good reason to *deny*, rather than to accept, Determinism. Thus, a "hidden line" – in the sense of a sharp boundary – *between truth and falsity*, is not something to which one is committed merely by being a Bivalentist and embracing classical logic. If SBT were true, a Classical Bivalentist might still coherently deny that there is a sharp boundary *between truth and falsity*. S/he might, *e.g.*, claim only that there is a sharp boundary between determinate truth and indeterminate truth. *I* do not wish to claim this; for I do not think we have good reason to believe SBT. It is worth pointing out, however, that some Non-Bivalentist deniers of Determinism (*e.g.* Burgess 1998) have defended the claim that higher-order vagueness might have an upper limit – which seems equivalent to the claim that there might be a sharp boundary between predications that are "perfectly determinate", and those for which there is "some order" of indeterminacy.

I have two merely speculative suggestions to make in response to Burgess. I do not propose to defend these suggestions.

My first speculative suggestion is this. Perhaps an Anti-Determinist embracer of SBT might say that the sharp boundaries lie at either end of the sorites series, so that there is *some* order of indeterminacy, however minuscule, at *any* predication of a vague predicate, even at 'Hairless George is bald'.

My second speculative suggestion in response to Burgess is this. Perhaps higher-order vagueness terminates at some order, but it is indeterminate which order it terminates at.

2.9 CONCLUSIONS

2.9.1 Some conclusions

Even Classical Bivalentists have good reason to deny Determinism. Borderline predications of vague predicates are indeterminate. In particular, given the arguments in Chapter 1, *unless* the Simple View both is true and is an effective way of resisting the Combined Spectrum Argument, there may be indeterminate statements of diachronic personal identity.

There is no reason to think that sorites series have sharp boundaries at all, let alone that the boundary between truth and falsity must be sharp.⁵³

2.9.2 A further conclusion

We can account adequately for vagueness and indeterminacy without rejecting Bivalentism or classical logic. Given the comparative simplicity and familiarity of Bivalentism and classical logic, we ought not to reject them without good reason. We ought therefore to be Anti-Determinist Bivalentists, and to accept classical logic – unless problems unrelated to those of vagueness and indeterminacy give us good reason to reject Bivalentism or classical logic.

One might take oneself to have good reason to reject the further conclusion just stated, yet still agree with the conclusions stated in Subsection 2.9.1.

2.9.3 Some further remarks

I suggest that the deniers of the epistemic view disagree over which logical rules to agree upon *using*, and that this dispute is therefore a practical one: Which logic *ought* we to use?

It seems to me preferable to retain at least classical rules of inference, while at the same time bearing at the forefront of our minds that the logic we use is one thing, while the question of *indeterminacy* is another.

⁵³ Alter and Rachels (2004) argue that even if the epistemic view is true, a variation of the Combined Spectrum Argument should still persuade us of a claim that is similar to Personal Indeterminism in spirit, *viz.* the claim that, though the facts would determine a sharp boundary in the Combined Spectrum, the *extra-linguistic* facts would not do so by themselves. I have been interested, however, in defending the strict version of Personal Indeterminism, according to which SDPIs can be indeterminate in a full-blooded sense. Along the way, I hope to have made sense of the general notion of (non-epistemic) indeterminacy.

That what logic we use is one thing, while whether vagueness gives rise to indeterminacy is another, is shown by the fact that the epistemic view is not a consequence of Classical Bivalentism. It is not plausible even given Classical Bivalentism.

This divide is evident, too, when we consider that the logical connectives themselves are subject to conventions. They are symbols, and they are in fact used in particular ways. In other words, conventions surrounding ‘or’, ‘not’ and other logical connectives might be one way, or they might be another.

Consider that, among those who have agreed to reject Excluded Middle for borderline predications, an assertion of ‘*S* or not-*S*’, where ‘*S*’ is known by the speaker to be a borderline predication, would be misleading. Everyone in such a community would assume that nobody follows a rule that says, where ‘*S*’ is a borderline predication, to believe ‘*S* or not-*S*’. In this community, saying ‘*S* or not-*S*’ is a way conveying information: it is a way of conveying that ‘*S*’ is not indeterminate.

Conversely, among those who have agreed upon classical conventions, assertions of the statement ‘*S* or not-*S*’ never convey factual information. This statement is on these conventions true regardless of the facts. Given such conventions, such an assertion, even if ‘*S*’ is a borderline predication, is never misleading. Some other means must be used to convey information regarding whether ‘*S*’ is indeterminate or determinate – for example, the expression ‘it is determinate that’.

But once we have agreed upon a particular set of logical conventions, regardless of whether they are classical or non-classical, there will always remain among users of such a common predicate as ‘tall’, ‘big’ or ‘thin’, a multiplicity of application rules for that predicate that are *not* ruled out. This situation is an unavoidable consequence of the practicalities of communication; it is impracticable for everyone to agree on a single precise rule for every single predicate; and no set of logical conventions could impinge upon this pragmatic state of affairs.

Chapter 3

THE SIMPLE VIEW

3.1 INTRODUCTION

3.1.1

I concluded in Chapter 1 that Personal Indeterminism is true unless either (a) the epistemic view of vagueness is correct, or (b) the Simple View of DPI both is true and is an effective way of resisting the Combined Spectrum Argument. Chapter 2 has ruled out (a). In the present chapter I shall address (b), arguing that the Simple View should be rejected.

The *epistemic* view is defended by only a few philosophers. Nevertheless, the arguments *for* the epistemic view pose a very significant challenge for any defender of genuine indeterminacy. In order comprehensively to meet this challenge, we had to offer an explication of the notion of indeterminacy. That explication involved, among other things, drawing the important distinction between two kinds of variation, *viz.* variation between possible worlds, and variation between different admissible precisifications. For one and the same world might be described in multiple ways. (See Section 2.7, especially Subsection 2.7.2.)

The Simple View, like the epistemic view, is defended nowadays by very few philosophers. Nevertheless, the arguments for the position warrant careful attention here. In addressing these arguments, I aim to do more than simply to defend Personal Indeterminism by showing that this unpopular view is false. I aim, also, to diagnose some of the confusions underlying the Simple View. In Subsection 3.4.3, in particular, we shall see how one of these confusions is the failure to distinguish between two kinds of variation: (a) variation between different possible situations; and (b) variation between different imaginative presentations or pictures of situations.

A common thread, then, unites distinct sources of resistance to Personal Indeterminism. Both the epistemic view and the Simple View involve a failure to distinguish factual variation (variation between worlds or situations) from variation between different ways of describing or presenting a single set of facts – whether linguistically or imaginatively.

In Subsections 3.1.2 and 3.1.3 I shall define the Simple View of DPI and explain how a proponent of the Simple View could resist the Combined Spectrum Argument. In Section 3.2, I shall argue that the Simple View is *prima facie* implausible and should be disbelieved in the absence of a good argument in its favour. I shall also address the question of whether the Simple View receives support from an argument that facts of diachronic personal identity can be apprehended directly in virtue of the “subjective diachronic unity of experience”. In Subsections 3.3.1–3.3.4 I shall describe Williams’s (1970) “twin thought experiments”, explaining how they have been thought by some to effect an argument for the Simple View. In Subsections 3.3.5–3.3.10 I shall discuss various ways in which that argument might be resisted. In Subsection 3.4.1 I shall describe a certain “broader argument” for the Simple View. In Subsections 3.4.2–3.4.4 I shall critique the “broader argument”, diagnosing in the process some of the confusions that seem to underlie belief in the Simple View.

In Section 3.5, I will, as an aside, discuss some further responses to Williams’s twin thought experiments, addressing, in particular, the claim that they show the method of thought experiments to be flawed, and the claim that they show the concept of personal identity to be incoherent. I will reject both of these claims.

3.1.2 The Simple View

A proponent of the Simple View (a “Simple Theorist”) claims that, given an SDPI $\langle P, Q, t, t^* \rangle$ a possible world w in which all of Q ’s and P ’s physical and psychological features (ignoring the possibility that DPI might itself be called a “feature”) are just the same as their features in a distinct possible world g may nevertheless be such that in w , $P=Q$, but in g , $P \neq Q$. The Simple View says even more than this, in fact: it asserts that worlds may differ in respect of DPI facts without differing in respect of facts of *any* other kind.

Note that denying the Simple View (*i.e.* defending the “Complex View”) does not entail claiming that the features upon which *whether* $P=Q$ supervenes are such as to *determine*, always, either that $P=Q$ or that $P \neq Q$. On Personal Indeterminism,

sometimes, they may leave it undetermined whether $P=Q$. But that is not to say that there is a further *fact* which could be one way or the other. Rather, what variability remains is as between different precisifications. On some precisifications of ' $P=Q$ ' (we would say that) $P=Q$. On others (we would say that) $P\neq Q$. So we could not without misleading others describe to *them* the facts either by saying simply ' $P=Q$ ', or by saying simply ' $P\neq Q$ '.

In contrast, the Simple Theorist says that, once all the particular psychological and physical facts are fixed, there remains variation as between possible worlds. There remains some further fact, of which we might be ignorant. That there is this "further fact" might be put thus: "the DPI facts do not just consist in non-DPI facts".⁵⁴

3.1.3 Resisting the Combined Spectrum Argument *via* the Simple View

Someone who holds the Simple View, and who wishes to deny Personal Indeterminism, might seek to undermine the Combined Spectrum Argument by showing that, if one has good reason to hold the Simple View, and one postulates that there is a sharp boundary in the Spectrum (one need not postulate where it is), that postulation is unassailable *via* the Combined Spectrum Argument. It is argued, in the following way, that the Combined Spectrum Argument would be question-begging, in trying to attack such a postulation.

The Combined Spectrum Argument relied for its force on the apparent slightness of the difference between every pair of neighbouring cases in the Spectrum. It seemed implausible to suppose that our concept of 'same person' is so precise as to draw a sharp boundary between two particular neighbouring cases, given the *slight* difference between these cases. But suppose that in every odd-numbered case, the person on the table after the operation is made to wear a gigantic hat; and in every even-numbered case, the post-operation person is made to wear no hat. In a sense, the difference between neighbouring cases would no longer be slight. But it would still be slight in all respects that are *relevant* to the question of whether the DPI predication is true. And it is only *relevant* slightness of difference between neighbouring cases that matters, if the Combined Spectrum Argument is to have force. Conversely, if there

⁵⁴ Johnston (1997: 262ff) argues that there are 'further facts of personal identity'. (264) But his understanding of what it is for a fact to be a 'further fact' does not commit him to the Simple View. For he still thinks DPI facts *supervene* on particular physical and psychological facts; and he thinks that as long as they so *supervene*, 'facts of personal identity may sometimes be indeterminate'. (265)

were a relevant respect in which two neighbouring cases did *not* differ slightly, but differed markedly, this might serve to undermine the argument, for one might argue that the alleged sharp boundary lies between these two cases.

The Combined Spectrum is described so that the relevant differences between neighbouring cases, in terms of all *physical* and *psychological* features, are slight. But what if facts about these features do not exhaust the facts that are relevant to whether the DPI predicate is satisfied? If the Simple View is true, then the fact \langle that $P=Q$ \rangle , or the fact \langle that $P\neq Q$ \rangle , does not simply consist in any set of other facts – be they physical or psychological facts. Rather, it is an irreducible *further* fact. Thus, an attempt to list all the facts that are *relevant* to whether DPI holds would arguably fail, if it included all the physical and psychological facts, but omitted this further fact: the DPI fact itself. Thus, though the Simple Theorist might regard particular physical *etc.* facts as *evidence* regarding whether $P=Q$, s/he must deny that these facts exhaust the set of facts that are relevant to whether $P=Q$. Conversely, the Complex Theorist can claim that the DPI fact itself is just entailed by, or is in some other sense already contained in, the physical and psychological facts; and so a list merely of all *these* facts contains all that is relevant to whether DPI holds.

It is now clear how a Simple Theorist might resist the Combined Spectrum Argument. The identity of buildings over time consists in more particular physical facts; and thus varying all these facts across a hypothetical spectrum, by degrees, challenges one who would deny the possibility of indeterminate building-identity statements to defend the implausible thesis that ‘same building’ is so precise a concept as to draw a sharp boundary between two cases that differ seemingly trivially with respect to all of these physical facts, which seemingly exhaust the facts that are relevant to whether a particular building is identical to a particular building. But the Simple Theorist claims that the identity of persons over time does not simply consist in other facts, but is itself a further fact, over and above any other facts. Thus two cases which differ but trivially with respect to all the particular psychological, physical *etc.* facts for which variation is gradual in the Combined Spectrum, may yet differ with respect to this further fact. To make the Combined Spectrum Argument persuasive for the Simple Theorist, one would have to make variation across the spectrum gradual with respect to the further fact of DPI; but to suppose that variation with respect to *this* fact can ever be anything other than abrupt – to suppose that it admits of degrees – is surely to beg the question against the denier of Personal

Indeterminism. If the Simple Theorist now asserts that there is a sharp boundary, between those operations where the further DPI fact ⟨that P is the post-operation person⟩ obtains, and those operations where the further DPI fact ⟨that P is not the post-operation person⟩ obtains (or, alternatively, those operations where the first, positive fact does not obtain), then this assertion of a sharp boundary is unassailable by the Combined Spectrum Argument, on pain of begging the question against the denial of Personal Indeterminism.

3.2 SOME INITIAL ARGUMENTS

3.2.1 The *prima facie* implausibility of the Simple View

Let us now ask, Which is more *prima facie* plausible: the Simple View or the Complex View?

Let person x exist at time t and person y exist at another time t^* . The Simple Theorist thinks that possible worlds g and w may be such that all the facts about the memories, intentions, physical qualities *etc.* of x and y are exactly the same – in fact *all* the facts are exactly the same – except for the bare fact that $x=y$ in g and $x\neq y$ in w .

The Simple View seems implausible – especially when we consider that the utterances by and beliefs of x , of y , and of everyone else, including those utterances and beliefs pertaining to whether x is y , are exactly the same in g as in w . The Simple View seems ontologically extravagant in positing a further fact despite this.

It may be thought that this kind of reasoning is verificationist. (See, *e.g.*, Swinburne 1974.) But it is merely Occamist. We need a *motivation* for positing the existence of this further, apparently undetectable bare fact – as Madell (1981: 10) acknowledges.

Swinburne (*e.g.* 1984: 41) thinks that memories and so forth may be *evidence* for DPI facts, and so, he seems to be claiming, this assertion of undetectability is false. But that argument, if Swinburne makes it, would be directed at the wrong kind of claim. We do not demand, from the Simple Theorist, an explanation of the kinds of things that could provide evidence for or against the holding of particular DPI facts. Rather, we want evidence for the holding of the Simple View. That is, we want evidence for the claim that DPI facts are *bare* facts. For the posit that they are not bare facts is more parsimonious, and, at least on an initial consideration, seems to leave

nothing unexplained. It is more parsimonious because it says that where we have all the facts about memories, physical continuity *etc.*, we have *all* the facts. But the Simple Theorist posits a *further* fact, the DPI fact itself.

Similarly, suppose I claim that once we have all the particular facts about building materials and their configuration, we have all the facts that are relevant to whether a building at one time is identical to a building at another time. I claim that facts of building identity are entailed by these particular physical facts. If you now disagree with me, and state that there is a *further* fact, a bare fact of building identity, which is not entailed by the physical facts about building materials and their configuration, your view of the diachronic identity of buildings is less ontologically parsimonious than mine. Accordingly, given that we have no reason to believe in a further fact of building identity, it unreasonable to believe that there is such a further fact. We ought to adopt the ontologically more parsimonious view. Similarly, unless someone can provide us with a special reason to think there are further facts of diachronic personal identity, we ought to disbelieve that there are these further facts.

3.2.2 Direct apprehension of DPI facts?

It might be retorted that we can apprehend DPI facts *directly*. The thought would be that our ability to be directly aware of a DPI fact, rather than aware only *via* memories, physical appearance *etc.*, is evidence that the DPI fact is a further fact over and above facts about memories, physical appearance *etc.*. Swinburne (1984: 42-44) thinks that (what I call) the subjective diachronic unity of one's experience (SDUE) is one's being directly aware of one's continued existence.

Consider that when I simultaneously hear a drumming noise and see a car, the noise and the car are felt as co-experienced: as experienced together *at one time* by a single subject. This is the "subjective synchronic unity of experience". Of course, this is not what the Simple Theorist about DPI needs, since it pertains to "synchronic identity", not to diachronic identity.

But there seems also to be a "subjective diachronic unity of experience" (SDUE). Consider this example of Swinburne's.

When a train moves along a railway line, the observer S on the bank has the following successive experiences: S sees train T at place p followed by T at place q; S sees T at q followed by T at r; S sees T at r followed by T at u, and so on. But then

that is not quite a full description of his experience. For if those were all the data of experience, S would need to *infer* [my emphasis] that the second experience which I have described succeeded the first (rather than being one which occurred on an entirely different occasion). Why he does not need to infer this is because it is itself also a datum of experience; S experiences his experiences as overlapping as in a stream of awareness. (1984: 43)

Often we can *infer* the temporal ordering of our experiences: 'Knowing that in general babies get bigger and not smaller I may infer that my seeing the small baby John occurred before my seeing the medium sized baby John, which occurred before my seeing the large baby John.' (Swinburne 1984: 42) But Swinburne argues that 'not all knowledge of the order of our experiences can derive from inference.' (42-43) I cannot infer that birds will always fly from left to right, or always from right to left. So

'[m]y knowledge that on a particular occasion my experience of seeing a bird on my left preceded my experience of seeing it on my right could not be an inference from the normal behaviour of birds... [S]econdly... [k]nowledge that events of type A normally precede events of type B must be grounded in many observations made by ourselves or others of actual such successions. (43)

The train example suggests that, just as we experience directly as co-experienced (by one subject) different simultaneous experiences, we can experience different experiences as co-experienced but also as experienced at different times. Arguably it would be a distortion of the concept of memory to say that we *remember*, in the situation Swinburne describes, the position of the train a fraction of a second ago. It is not obvious that memory accounts for SDUE.

Swinburne apparently characterises SDUE as direct awareness of DPI itself: '[T]he continuing existence of a person over a very short period of time is something which can often be experienced by that person.' (1984: 42)

Shoemaker (1984: 148-149) has argued *contra* Swinburne that SDUE is not direct awareness of DPI, since it is fallible, yielding false belief in case of fission.⁵⁵

⁵⁵ Actually Shoemaker (148) claims that SDUE is not *knowledge* of DPI unless coupled with the (surely actually safe) presumption that there has been no fission or fusion between the "co-
(footnote continued next page)

(See Chapter 4.) (Slightly question-beggingly: I will *fission* just in case each of two distinct future persons will be both psychologically and physically continuous with me-now.) Having fissioned half-way through a one-second interval j , the experiences had in the first half of j may still be felt as “co-experienced” with the experiences had during the second half of j , even though, *at least according to Shoemaker*, fission would have caused a change in identity of the experiencer. But this response to Swinburne depends on a particular theory about who is who in fission cases.

A better response is to contest the claim that the ability to be directly aware of one’s continued existence would count as evidence that the fact of one’s continued existence does not consist in (supervene on) other facts.

Consider pain. It seems I can be directly aware of the fact that I am in pain. I seemingly do not *infer* that I am in pain. But, as Garrett (1998: 116) points out, this direct awareness in itself does not preclude my pain’s supervening on other facts. Very plausibly, I cannot be in pain without certain physical features of my brain and body being a certain way. My being in pain would supervene on those features. But if I am directly aware of my pain, that direct awareness does not constitute good reason to think that pain supervenes on nothing else.

Similarly, if I am directly aware that the person x experiencing e is identical with the person y experiencing e^* , there is no obvious reason why this should preclude the supervenience of the fact $\langle \text{that } x=y \rangle$ on other facts. Certainly, there is no reason to think that if there is some kind of fact J , which supervenes on facts of another kind K , there is no requirement that knowledge of J -facts can be had only *via* knowledge of K -facts. Facts about the existence of water in certain places supervene on facts about the existence of H₂O molecules in certain places. But people knew plenty about water before they knew anything about molecules.

So even if we *never* knew anything about DPI *via* memories, intentions, physical appearance or *etc.*, this might still not be good reason, by itself, to rule out supervenience of DPI facts upon memories *etc.*⁵⁶

experienced” experiences. But this seems to be a stronger claim than Shoemaker needs to make, and seems a rather narrow interpretation of ‘knowledge’.

⁵⁶ Garrett: ‘We should distinguish the question of how we arrive at various personal identity judgements ... from the question of what makes such judgements true ... It is quite consistent to suppose that access to certain first-person judgements is ungrounded or criterionless, even though the truth-maker for such judgements involves essential reference to physical and psychological continuities (that is, criteria of personal identity).’ (1998: 116)

But in any case, we *do* draw inferences about DPI facts from facts about memories, physical appearance *etc.*. In ordinary life, over durations longer than a few seconds, we seem to answer questions about personal identity by consulting facts about physical or psychological features, all of which seem to admit of degrees of variability. We apparently do not have any kind of unmediated epistemic access to DPI facts not spanning a very short interval (and even in the latter case, it is highly debatable whether SDUE does yield such unmediated epistemic access). As Parfit (1984: 223-225) argues, when we are aware of our being identical to a particular person in the past, this awareness cannot (over any period of time of substantial duration) be distinguished from our awareness of particular physical and psychological features, such as memories *etc.*. Given an SDPI $\langle P, Q, t, t^* \rangle$, we judge whether $P=Q$ by examining particular physical and psychological features, just as, in judging whether a building existent at one time is identical to a building existent at another, we examine particular physical features, such as the composition of building materials, constituting physical continuity.

In summary, there seems to be no evidence for the Simple View, and plenty of evidence that physical and/or psychological continuity have some fairly significant role in relation to the continuing existence of persons. Therefore we should disbelieve the Simple View in the absence of some fairly persuasive argument in its favour. I shall consider shortly whether the “twin thought experiments” of Williams (1970) yield or suggest such an argument. Then I shall consider a certain “broader” argument for the Simple View.

3.2.3 The Simple View is not antireductionism

Note, incidentally, that arguments merely for *antireductionism* about DPI are not arguments for the Simple View. Antireductionists about DPI claim that we cannot give a reductive (non-circular) criterion for DPI. If no non-circular criterion of diachronic personal identity is plausible, this concerns the meaning of ‘same person’, ‘person’ *etc.*. But it does not concern the *ontological* nature of the *facts* of personal identity. Thus, antireductionism is consistent with the supervenience of DPI facts on other facts, and hence with the falsity of the Simple View.

3.3 WILLIAMS'S TWIN THOUGHT EXPERIMENTS

3.3.1

Williams (1970) poses a pair of thought experiments – or more accurately, two presentations of the same thought experiment – which I shall call his ‘twin thought experiments’, which have been thought by some, *e.g.* Madell (1981: 97-101), to constitute a convincing argument in favour of the Simple View, even though Williams himself uses them to argue for a physical criterion of personal identity. I shall call the argument for the Simple View that is based on Williams’s (1970) twin thought experiments ‘the Twin Experiments Argument’.

3.3.2 Williams’s twin thought experiments: part one

Suppose that we have two persons, *A* and *B*. Williams invites us to suppose that

it [is] possible to extract information from a man’s brain and store it in a device while his brain [is] repaired, or even renewed, the information then being replaced: it would seem exaggerated to insist that the resultant man could not possibly have the memories he had before the operation. ... [Imagine that] information [is] extracted into such devices from *A*’s and *B*’s brains and replaced in the other brain (Williams 1970: 162)

Williams (162) apparently intends this process to have a result as close to a physical *brain swap* as one can get, without the “mess” of actually swapping brains. The process might be said to involve “complete brain reconfiguration”.

Williams (1970: 163-167) presents an argument (which I shall call ‘Argument One’) (which argument he *later rejects*) that this information-swap between the brain of *A* and the brain of *B* constitutes a swapping by two persons of their *bodies*. Following Williams (163), let us use the term ‘the *A*-body-person’ to denote that person who, after the information-swap, inhabits the body which, prior to the information swap, belonged to *A*; and similarly for ‘the *B*-body-person’. Argument One, then, argues, in the following way, that *A* is the *B*-body-person, and *B* is the *A*-body-person.

Suppose we had told *A* and *B*, prior to the operation, that after the operation, one of the post-operation persons will be tortured and the other will be given a large

sum of money. 'We then ask each [of] *A* and *B* to choose which treatment should be dealt out to which of the persons who will emerge from the experiment, the choice to be made (if it can be) on selfish grounds.' (Williams 1970: 163)

If *A* chooses in favour of the *B*-body-person (*i.e.* chooses for the *B*-body-person to get the money and the *A*-body-person to get the torture), and *B* chooses in favour of the *A*-body-person, and the *B*-body-person is tortured, then the *B*-body-person

will not only complain of the unpleasant treatment as such, but will complain (since he has *A*'s memories) that that was not the outcome he chose, since he chose that the *B*-body-person should be well treated; and since *A* made his choice in selfish spirit, he may add that he precisely chose in that way because he did not want the unpleasant things to happen to *him*. The *A*-body-person meanwhile will express satisfaction both at the receipt of the \$100,000, and that the experimenter has chosen to act in the way that he, *B*, so wisely chose. (Williams 1970: 164)

Williams argues that this is a 'strong case' for saying that the experimenter, in favouring the *A*-body-person, causes *B*, and not *A*, to get the outcome he wants. 'It is therefore a strong case for saying that the *B*-body-person really is *A*, and the *A*-body-person really is *B*'. (164) This hypothetical case, then, seems naturally to be construable as one of persons swapping bodies; and thus it supports the thesis that physical continuity is not necessary for personal identity.

Williams reinforces the point by considering other combinations of choices by the pre-operation persons. (165-166) For example, suppose *A* chooses in favour of the *A*-body-person, and *B* chooses in favour of the *B*-body-person. Suppose after the operation the *A*-body-person gets the money and the *B*-body-person gets the pain. Williams argues that '[b]oth the *A*-body-person and the *B*-body-person will have to agree that what is happening is in accordance with the preference that *A* originally expressed.' But while '[t]he *B*-body-person will naturally express this acknowledgment (since he had *A*'s memories) by saying that this is the distribution he chose... he (the *B*-body-person) ... does not like what is now happening to him'. On the other hand, '[t]he *A*-body-person will recall choosing an outcome other than this one, but will reckon it good luck that the experimenter did not do what he recalls choosing.' It seems natural, then, to think that although 'the *A*-body-person has gotten

what he wanted' he did not get 'what he chose'. And it looks as though 'the *B*-body-person has gotten what he chose, but not what he wanted.' So it seems, again, that the *A*-body-person is *B* and the *B*-body-person is *A* – again, a body-swap – but in this case, 'the original choices of both *A* and *B* were unwise.' Or so goes Argument One. (Williams 1970: 165)

3.3.3 Williams's twin thought experiments: part two

In the second of his twin thought experiments, Williams (1970: 172) asks us to imagine the following spectrum of hypothetical operations any one of which might be performed on *A*, and that after any such operation, the person (whoever he is) then inhabiting the body that was *A*'s body before the operation will be tortured. Before any such operation is performed, *A* would be informed of all the details of the operations and of the subsequent torture.

- (i) *A* is subjected to an operation which produces total amnesia;
- (ii) amnesia is produced in *A*, and other interference leads to certain changes in his character;
- (iii) changes in his character are produced, and at the same time certain illusory "memory" beliefs are induced in him; these are of a quite fictitious kind and do not fit the life of any actual person;
- (iv) the same as (iii), except that both the character traits and the "memory" impressions are designed to be appropriate to another actual person, *B*;
- (v) the same as (iv), except that the result is produced by putting the information into *A* from the brain of *B*, by a method which leaves *B* the same as he was before;
- (vi) the same happens to *A* as in (v), but *B* is not left the same, since a similar operation is conducted in the reverse direction.

(Williams 1970: 172)

Call the above Spectrum of cases ‘Williams’s Spectrum’. Clearly case (vi) is just the first of the twin thought experiments, described in a different way: it is an information-swap between brains. Does personal identity “follow” the information, or does it “follow” the physically continuous bodies and brains? Argument One suggested that it follows the information. But, on the basis of the above presentation, Williams (1970: 172-180) argues, in what I shall call ‘Argument Two’ (see below) that it follows the bodies.

Now it might be argued (see *e.g.* Beck 1998 §6) that Williams (1970: 172) describes the second of his twin thought experiments in a question-begging or at least a leading way. *E.g.*, Williams writes: ‘*A* is informed beforehand that just these things [the things constituting one of the operations] followed by the torture will happen to *him*.’ (My emphasis.) I shall ignore this apparent question-beggingness though; arguably whatever rhetorical help it offers Williams’s case is shallow enough that Williams’s argument would still have force were the description more neutral. And the remaining force can in any case be dealt with, as we shall see. Furthermore, Williams consciously argues that this question-begging description is just the correct one. (See 1970: 169.)

Argument Two may be summarised as follows.

- (Step 1) *A* is the *A*-body-person in (*i*).

- (Step 2) Given that *A* is the *A*-body-person in (*i*), he is also the *A*-body-person in (*ii*).

- (Step 3) Given that *A* is the *A*-body-person in (*ii*), he is the *A*-body-person in (*iii*).

- (Step 4) Given that *A* is the *A*-body-person in (*iii*), he is the *A*-body-person in (*iv*).

- (Step 5) Given that *A* is the *A*-body-person in (*iv*), he is the *A*-body-person in (*v*).

(Step 6) Given that *A* is the *A*-body-person in (v), he is the *A*-body-person in (vi).

(Conclusion) *A* is the *A*-body-person in (vi).

Williams (1970: 172-174) argues for each of Steps 1 to 6 individually. Williams apparently assumes that *A*, before the operation, would be rational in fearing the torture of the *A*-body-person if and only if *A* is the *A*-body-person. He seems, accordingly, to use the rationality or otherwise of a person *x*'s fear of a person *y*'s torture as a guide to the question of whether *x* is diachronically identical to *y*. Williams argues that in (i) *A* has reason to fear the torture: amnesia does not reduce the fearfulness of this prospect. (172) In (ii) and in (iii) there is, argues Williams, no reduction in the rationality of *A*'s fearing the torture merely on account of the psychological changes preceding it. The fact that the memories to be implanted in the *A*-brain (*i.e.* the brain which, before the procedure, is *A*'s) happen to have a model in (iv) is not, argues Williams, a significant difference between this case and (iii). (172) Williams now argues that, in (v), '[i]f we concentrate on *A* and the *A*-body-person, we do not seem to have added [in (v)] anything which from the point of view of his fears makes any material difference.' (173) To *A*, 'looking forward', it does not make any difference whether the implanted memories (or apparent memories) will be transferred directly from the *B*-brain or merely modelled after those in the *B*-brain. Furthermore, adds Williams, we have in (v) an 'undisputed *B* in addition to the *A*-body-person', so the *A*-body-person is not *B* – making it hard to say who *else* he could be but *A*. (173) Finally, argues Williams, since the *A*-body-person in (v) is 'in character, history, everything, exactly the same as the *A*-body-person in (vi)', the *A*-body-person is *A* in (v) only if he is *A* in (vi). So since he is *A* in (v), he is *A* in (vi). (174)

Williams (174-179) says more besides. But before going on (in Subsection 3.3.10) to consider what else Williams says, let us see in what sense Arguments One and Two, taken together, might be thought to effect an argument for the Simple View; and let us examine some ways in which this argument might be resisted.

3.3.4 The Twin Experiments Argument for the Simple View

Madell, a Simple Theorist, thinks that, in presenting to us the twin thought experiments, '[w]hat Williams has ... done, without recognising it, is to tell two stories which together powerfully support the ... Cartesian view of the self'. (1981: 98) (For current purposes, 'the Cartesian view' is interchangeable with 'the Simple View'.)

Given any particular performance of operation (vi), the conclusion of Argument One and the conclusion of Argument Two cannot both be right. A cannot be the B -body-person *and* the A -body-person. But the Simple Theorist might argue that nevertheless there is nothing especially implausible about either argument. S/he might argue that the plausibility of each argument is indicative of the possibility of our surviving *any* kind of change, psychological or physical. According to this line of reasoning, on any particular performance of operation (vi), whether A is the A -body-person or the B -body-person cannot be inferred merely from an account of particular physical and psychological events. This shows that whether A is the A -body-person or the B -body-person is *independent* of facts about these events: it does not supervene on them. For Madell (1981: 99), the twin thought experiments '[lead us] to understand just how it is possible for there to be these two different possibilities, and for there to be no observable difference between them.' That is, by showing us the coherence *both* of A 's surviving as the A -body, *and* of A 's surviving as the B -body, the thought experiments show us that, given the *same* set of "non-DPI facts", there are two *distinct possibilities* regarding the "DPI facts". So possible worlds may differ *only* in respect of "DPI facts"; so the Simple View is true. Or so goes the Twin Experiments Argument for the Simple View.

There would appear, at least on an initial examination, to be two ways of responding to the twin thought experiments without yielding to the Simple View: one might attack Argument One; or one might attack Argument Two. These ways may not be mutually exclusive. Noonan, for example, has attacked both arguments, as we shall see. There are other ways of responding to the argument; but we shall come to these later.

3.3.5 Resisting Argument One

This is Williams's own preferred option. Williams finds Argument Two persuasive, and concludes that *A* is the *A*-body-person. He seems less than certain that this is the correct response (1970: 179), but thinks it better than the alternative. The persuasiveness of the first presentation (*i.e.* the presentation of operation (*vi*) as a body-swap, as under Argument One), argues Williams, 'turned on the extreme neatness of the situation in satisfying, if any could, the description of "changing bodies." ' (179) This 'neatness', however, is contrived, a 'product of the will of the experimenter to produce a situation which would naturally elicit, with minimum hesitation, that description.' (179) Had any one of many slightly different operations, *e.g.* (*v*), been performed, we would not tend to think of the result as a body-swap. 'The experimenter', argues Williams, 'has not... *induced* a change of bodies; he has rather produced the one situation out of a range of equally possible situations which we should be most disposed to call a change of bodies.' In comparison, 'the principle that one's fears can extend to future pain whatever psychological changes precede it seems positively straightforward.' (Williams 1970: 180)

Noonan also suggests that Argument One can be resisted, but on different grounds; he certainly is not persuaded by Argument Two. Noonan (1989: 181-185) argues that Argument One, if persuasive, is so only because Williams ignores that the choices made by *A* and *B* before the operation, and their reflections after it, would depend crucially on *A*'s and *B*'s views on what is the correct criterion of personal identity.

I will not re-iterate all of Noonan's response to Argument One here, but to give the gist of it, here is what he says about the case in which, when asked which post-operation person should be tortured, and which rewarded, *A* and *B* both answer in accordance with their belief in the bodily criterion of personal identity. *A* believes he will be the *A*-body-person, and chooses in favour of the *A*-body-person; and *B* believes he will be the *B*-body-person and chooses in favour of the latter. After the operation, the *B*-body-person is tortured, and the *A*-body-person is given the money. Noonan agrees with Williams that in this case both post-operation persons will acknowledge that the outcome is in accord with the preference *A* expressed. But Noonan thinks Williams wrong in thinking 'the *B*-body person will 'naturally express' his acknowledgement [of this] by saying that this is the distribution he chose.' (Noonan 1989: 183) Noonan thinks the *B*-body-person 'will merely have a

tendency to express this acknowledgement in this way [on account of his memories or apparent memories] ... a tendency he will try to suppress: his considered statement will be to the effect that the distribution is not the one he chose, but merely the one he has an illusory 'memory'-impression of choosing.' (183) Similarly, the *A*-body-person will *not* think himself fortunate not to have had *his* wishes abided by, but will rather think it fortunate merely that the wishes he has the *illusion of remembering* were not abided by.

Noonan considers other combinations of choices and attitudes by *A* and *B*, thereby arguing that Williams's presentation of Argument One relies for its persuasiveness on ignoring the possibility of the subjects' believing in the bodily criterion and forming post-operative responses on the basis of that belief.

One might in the end have reasons for thinking the conclusion of Argument One correct; for one might have independent reasons to believe in a criterion of personal identity that does not require physical continuity at all, but takes psychological continuity to be a sufficient condition of personal identity. But Noonan's point is that the story told by Williams is not especially persuasive, in itself, that (vi) is a body-swap. Indeed, as we shall see, Noonan *also* finds fault with Argument Two. For Noonan, one seeking to resist the Twin Experiments Argument for the Simple View is spoilt for choice.

Noonan's response to Argument One has been criticised by Beck (1998 §3), who thinks it 'fairly uncontroversial' that even firm believers in the physical criterion 'will take themselves to be the persons they feel like, and whose lives they remember.' For Beck,

it certainly does not follow from the fact that *A* overtly adopts the physical criterion that if *A* were to look in the mirror and see *B*'s body, he would say, 'Oh look, I'm not the person I think I am!' Indeed, it is most implausible that *A* would react in this way, yet that is precisely what Noonan is suggesting. (1998 §3)

Although Beck, in a sense, misses Noonan's point, Noonan's critique of Argument One remains unconvincing.

Here is the sense in which Beck misses Noonan's point. Let us suppose that the *A*-body-person has a beard and the *B*-body-person is clean-shaven. Were the *B*-body-person, who has *A*'s memories and *A*'s beliefs about personal identity, to look in

the mirror, the *B*-body-person would, if he conscientiously believes in the physical or bodily criterion of DPI, think, ‘Since the bodily criterion is true, these apparent memories I have are false memories: I seem to remember having a beard; but this memory is illusory, because, as I know from the bodily criterion, *I* was whoever’s body *this* (the *B*-body) is all along, and *this* body does not have a beard (unless that scientist shaved it during the operation).’ So, *contra* Beck, the *B*-body-person (whom I presume he means his ‘*A*’ to denote) would *not* think ‘I’m not the person I think I am’, or at least, not on a considered reflection. For Noonan admits that he would *tend* to think this; but he would suppress this tendency on account of his belief in the bodily criterion. So in this sense, Beck’s critique misses Noonan’s point.

But here is why Noonan’s critique of Argument One remains unconvincing. Argument One does not purport to *prove* (deductively) that *A* is the *B*-body-person. But it seeks to persuade us that the scenario is more plausibly regarded as a body-swap than as anything else. And although the fact that the resulting persons’ considered statements and beliefs about their identities may not invariably support the psychological criterion, the great mental effort which would seem to be required on the part of the *B*-body-person if he is to think of himself as *not* having undergone a body-swap is indicative of the naturalness of regarding the operation as effecting a body-swap.

3.3.6 Resisting Argument Two

These days, the most popular response to the twin thought experiments seems to be to resist Argument Two. Different authors have found different faults with it. Critiques of Argument Two may be divided into three types: (a) arguments that a sharp line may reasonably be drawn between a particular pair of cases in Williams’s Spectrum; (b) arguments that a range of points in Williams’s Spectrum might be reasonable locations for a sharp boundary, depending on how certain underdescribed aspects of the cases are “filled out”; and (c) accusations of soriticality⁵⁷.

⁵⁷ I say that something is soritical if it has the form of a sorites argument, but is not a valid mathematical induction. I say that an argument “has the form of a sorites argument” even if its conclusion is not *obviously* false. The form of a sorites argument is exemplified by the Tadpole Paradox (see Subsection 2.2.1).

3.3.7 Drawing a line

Argument Two has six steps, each of the form: 'Given that *A* is the *A*-body-person in (case *n*), he is the *A*-body-person in (case *n*+1).' If one can argue convincingly of a particular *n* that it is reasonable to think *A* is the *A*-body-person in case *n* but *not* in case *n*+1, then one can undermine Argument Two.

Noonan (1989: 185-190) argues that a sharp line might reasonably be drawn between (iv) and (v), because in (v), not only do 'the *A*-body person's memory-impressions have a model which is also their cause', but also, we may infer that a 'very special causal process' (189) ensures that the information transfer is preceded by a complete and *irrevocable* psychological erasure; whereas in (iv) the psychological changes might not be irrevocable. For a reason for supposing persons to continue to exist in spite of *amnesia* is that the amnesia may not be permanent: the memories might eventually be recovered; and in case (iv), if all that happens insofar as the *deletion* of *A*-brain's previous memories is concerned is *mere* amnesia, then this is consistent with the potential recoverability of the deleted memories, and so with the reasonable identification of *A* with the *A*-body-person in (iv). But in (v), the deletion is presumably irrevocable, since the information-transfer process envisaged in (v) and (vi) (and in the first of the twin thought experiments) has to effect as irrevocable a deletion of memories from the skull as would be effected by the physical removal of the entire brain: 'It must involve [irrevocable deletion] otherwise the psychological theorist would not have to accept (vi) as a clear case of bodily interchange', argues Noonan. (189) This point enables one to resist Williams's Argument Two while retaining belief in 'common-sense facts about survival through amnesia (ordinarily so-called), character change and inducement of illusory memory impressions.' (189) And it enables Noonan to rebut Williams's (1970: 173) claim that the difference between (iv) and (v) is merely a difference in the *provenance* of *A*-brain's new memories.

Noonan also suggests that, for some theorists, *B* might reasonably be regarded as undergoing *fission* in (v) but not in (iv), making for a further significant difference between these cases. (189) I shall discuss the notion of fission in Chapter 4; but the important thing to note here is that, if *B* undergoes fission in (v), it is not the case that, in (v), we have, as Williams (173) claims, an '*undisputed B* in addition to the *A*-body-person' (my emphasis). (Noonan 1989: 189)

So much for Noonan's critique of Argument Two.

Beck (1998 §6) suggests that, given Williams's description of Williams's Spectrum, we can infer that the difference between cases (ii) and (iii) is much more significant than what it appears superficially to be, suggesting that a sharp boundary might be drawn there. Consider that in case (iv), the changes are such that the resulting totality of the psychological characteristics of the post-operation person is qualitatively just like the totality of the psychological characteristics of another actual person, *B*. This must be so, given that the changes in (v) and (vi) yield the same post-operative totality of psychological characteristics. But then case (iv) is supposed to differ from (iii) *only* in that, in (iv), there exists an actual other person the totality of whose psychological characteristics the post-operative patient's post-operative totality of psychological characteristics happens to match. So given that (ii) differs from (i) merely in that, as well as amnesia, there are in (ii) 'certain changes in ... character' (Williams 1970: 172), the jump from (ii) to (iii) must actually be quite radical, despite the tendency of Williams's description to suggest otherwise; for 'the character and memory changes occurring in (iii) are sufficiently different from A's to be those of a totally different, even if non-actual, person. ... Were the extent of the change in (iii) made clear in the description... it is much less likely that anyone would find the move from (ii) to (iii) acceptable.' (Beck 1998 §6)

In fact, while Beck thinks this a reasonable response to Argument Two in its original formulation, he thinks the argument might be revised so as to evade this response – but that in its revised form, it succumbs to a different objection – as we shall see in Subsection 3.3.9.

3.3.8 Arguing that the cases are underdescribed

While Noonan suggests that a line might reasonably be drawn between, in particular, (iv) and (v), and Beck prefers a line between (ii) and (iii), Garrett argues that '[f]or all that has been said [by Williams], a line can defensibly be drawn anywhere between (i) and (vi)', depending on how certain 'under-described factor[s]' in the thought experiment, *e.g.* 'extent and type of memory loss', are filled out. (Garrett 1998: 54) Garrett argues that if the memories erased in operation (i) are only memories of *A*'s *experiences* (as opposed to memories of factual knowledge and of abilities), then plausibly *A* survives through (i); but depending on how drastic the psychological changes produced in (ii) or (iii), we could reasonably regard *A* as perishing in either of these operations. (54) This response seems in the spirit of Beck's: Williams's scanty

descriptions of the changes in each operation gloss over what must in fact be quite extensive changes, given what we can infer *must* happen, if we are to get from (i) to (vi).

We might add that, if Noonan is right about the relevance of the potential recoverability, or irrecoverability, of the erased memories, another respect in which Williams's Spectrum is significantly under-described is in not specifying which operation is the first in the Spectrum to effect irrevocable memory deletion.

3.3.9 Accusing Argument Two of soriticality

Having said this, drawing a boundary at a particular point in Williams's Spectrum, even if we were given a more thorough description of each operation in it, is perhaps not a very satisfying response, for the following reasons.

Firstly, the variety of suggestions about where such a boundary might be drawn indicates that it is at least not obvious of any particular location that it is where the boundary lies. Perhaps it might become more obvious once a complete description is provided; but then perhaps it might not.

In any case, as Beck (1998 §3) has suggested, one might revise Argument Two by describing more possible operations intermediate in nature between the operations in Williams's original Spectrum – a range of cases exhibiting similar gradualness of variation as is exhibited in Parfit's (1984: 231-233) Psychological Spectrum – but in which Williams's operation (i) is at one end of the spectrum and his operation (vi) is at the other end. This "Revised Spectrum" would make the drawing of a sharp boundary anywhere in the Spectrum appear unreasonable – as unreasonable as it appeared in the Combined Spectrum.

But of course, as Beck (1998 §6) and Garrett (1998: 54) point out, such a Revised Argument Two would have a familiar flaw: it would be a sorites argument. While one might agree with Williams that each of its steps is *plausible*, and no particular location is uniquely suited for the drawing of a boundary, one might claim that not all the Revised Argument's premisses (understood as material conditionals) are true, although it is indeterminate which one is untrue. This was a good response to the Tadpole Paradox (Section 2.2). Is it not then a good response to the structurally similar Revised Argument Two?

Alternatively, one might simply respond by saying that, even if we have not worked out what exactly is the correct diagnosis of sorites paradoxes, we know that

they are unsound, since their conclusions are obviously false. An argument of sorites form is unsound, then, and Argument Two, or at least Revised Argument Two, has this form.

The important point is that, when confronted with a sorites argument for a conclusion we have good reason to reject, we are not required, in order to defend the contradictory of that conclusion, to justify the denial of any particular conditional premiss in the sorites argument. We need only to point out that the argument form is soritical and hence is fallacious – just as the Tadpole Paradox is fallacious.

Note, however, that this response to Revised Argument Two would involve an appeal to the possibility of vagueness in statements of diachronic identity. For sorites arguments are unsound because they turn on the vagueness of linguistic expressions in their premisses. If a superficially sorites-like argument is constructed on the basis of a *precise* predicate – for example, the predicate ‘is an integer’, with respect to the series of integers – then there is no grounds for claiming that it is unsound. On the contrary, the argument would not really be soritical: it would be a valid mathematical induction.

3.3.10 Williams’s own argument against Personal Indeterminism

Williams (174) recognises the obvious similarity between Argument Two and a sorites argument. But Williams (1970: 174-179) argues – in what I shall call his ‘Argument Three’ – that if we refuse on sorites-related grounds to draw a sharp boundary, holding that one or more of the cases (*i*) to (*vi*) leaves it indeterminate whether *A* is the *A*-body-person, we are committed to something implausible. He argues, on the basis of the first-person perspective that is distinctive of personal identity, that it cannot be indeterminate whether one is identical to some person existing in the future.

We have examined Williams’s Argument Three already: it was just the argument we discussed in Section 1.6, under the heading ‘Williams’s Torture Case’. We saw that Noonan (1989: 190-195) has defeated Argument Three.

3.4 A BROADER ARGUMENT FOR THE SIMPLE VIEW

3.4.1

Perhaps a Simple Theorist would say that we have missed the point entirely. S/he might say that the soritical argument for the conclusion that *A* is the *A*-body-person is, in fact, a distraction, at best a rhetorical device for loosening the mindset of opponents of the Simple View. ‘Let us forget cases (*i*) to (*v*),’ s/he might say, ‘and focus just on (*vi*).’ And if we do so focus, there seems still to be some case to answer. For it seems I can *imagine* waking up with new memories just as easily, or almost as easily, as I can imagine waking up with a new body.

Madell (99-101) takes the twin thought experiments to be just a particularly forceful illustration of a broader point. Given a particular set of non-DPI facts, we can coherently imagine, together with these facts, multiple distinct possibilities with regard to our own identity, and we can imagine these possibilities *as* distinct (or so it is alleged). Although the possibilities may be indistinguishable from the point of view of an observer, the first-person viewpoint distinguishes them very clearly, argues Madell (1981: 100): ‘*A* can easily form a conception of how his experience would be different’ in each ‘story’ (the presentation of operation (*vi*) as under Argument One being one story and the presentation of operation (*vi*) as under Argument Two being another). In the second story, ‘instead of *A* having the experience of moving from body *A* to body *B*, he has the delusive experience of having moved from body *B* to body *A* (since a set of memory impressions which correspond to *B* are implanted into him). And this is surely a difference in experience of which *A* in the first story can form a clear conception.’ (100) Thus, though *A* in the first story undergoes one set of experiences, he could conceive how things would have been rather different for him had the second story turned out instead to have been the one that actually unfolded.

And Madell (1981: 100) claims that even if *A* could *not* imagine how his experience would be different in the two stories, this would not entail that they are not distinct possibilities.

Swinburne (1984) appeals to a similar line of reasoning. He argues that ‘it is coherent to suppose a person could continue to exist with an entirely new body’ or even ‘with no body at all’. He argues, furthermore, that a person could lose all memories and still exist: ‘Quite clearly, we do allow not merely the logical possibility, but the frequent actuality of amnesia – a person forgetting all or certain

stretches of his life.’ (24) It may be reasonable to think a person can survive psychological erasure if s/he keeps the same body, or to think a person survives a body swap if s/he keeps the same memories. But Swinburne argues that a person might survive radical physical *and* psychological change, and that, even though the true identity fact would in such a case be undetectable, ‘only given verificationist dogma, is there any reason to suppose that the only things which are true are those of whose truth we can have evidence’. (24) (Swinburne later argues against the relevant form of verificationism.)

Swinburne points out that ‘many religions have taken seriously stories of persons passing through the waters of Lethe [the memory-erasing river] ... and then acquiring a new body.’ Also, ‘[t]hose who hope to survive their death, despite the destruction of their body, will not necessarily be disturbed if they come to believe that they will then have no memory of their past life on Earth. ... [T]here seems to be no contradiction involved in their belief.’ (25) Such stories may not be believed: but disbelieving such stories is one thing, while holding them to be *contradictory* is quite another. Swinburne acknowledges that some stories contain contradictions that are hidden, ‘[b]ut’, he argues, ‘the fact that there seems (and to so many people) to be no contradiction hidden in [these particular stories] is good reason for supposing that there is no contradiction hidden in them – until a contradiction is revealed.’ (Swinburne 1984: 25)

Swinburne⁵⁸ argues that, furthermore, not even the laws of nature necessitate that ‘a person have a body made of certain matter, or have certain apparent memories, if he is to be the person which he is.’ (1984: 25) Argues Swinburne:

[L]et us assume that natural laws dictated the course of evolution and the emergence of consciousness. ... Natural laws then, we assume, dictated how this globe could evolve, and so which arrangements of matter will be the bodies of conscious men, and just how apparent memories of conscious men depend on their brain states. ... [but] what natural laws in no way determine is which ... body is yours and which is mine. Just the same arrangement of matter and just the same laws could have given to me the body (and so the apparent memories) which are now yours, and to you the body (and so the apparent memories) which are now mine. It needs either God or chance to

⁵⁸ For his argument for this claim, Swinburne credits Knox, J. Jr 1969. ‘Can the self survive the death of its mind?’ *Religious Studies* 5: 85-97.

allocate bodies to persons; the most that natural laws determine is that bodies of a certain construction are the bodies of some person or other ... Since the body which is presently yours (together with the associated apparent memories) could have been mine (logic and even natural laws allow), that shows that none of the matter of which my body is presently made (nor the apparent memories) is essential to my being the person I am. That must be determined by something else. (Swinburne 1984: 25-26)

Hence, so the argument goes, the particular physical and psychological facts fail to fix the facts about “who is who”, either logically or in virtue of the laws of nature.

For Madell and Swinburne, the Twin Experiments Argument just brings out this broader point: that we can coherently envisage surviving virtually any kind of change. The challenge is to say what is supposed to be *impossible* about what is thereby envisaged. To undermine Argument Two by accusing it of soriticality is to miss this broader point, of which William’s thought experiments are merely an illustration.

Williams himself – although he would not go nearly as far as Madell and Swinburne – seems to think that Argument Two, at least, is just a reinforcement of an already very plausible ‘principle’, *viz.* that ‘one’s fears can extend to future pain whatever psychological changes precede it’. (1970: 180) The Simple Theorist would claim that another principle is *also* true: ‘One’s fears can extend to future pain whatever *physical* changes precede it.’ And yet another principle: ‘One’s fears can extend to future pain *whatever* changes precede it.’ The thought is that it is rational for me to fear a future experience if and only if it is *I* who shall experience it. That our fear of pain can “reach through” any kind of change is just indicative of the fact that DPI facts do not supervene on the physical and psychological facts that would specify the various kinds of “changes” – or so the argument goes.

3.4.2 Responding to the broader argument (I): transworld identity distinguished from diachronic identity

An initial point to make in response to at least some of Swinburne’s and Madell’s points is that *transworld identity* is to be distinguished from *diachronic identity*. This is a point made by Garrett (1998: 30).

To say ‘I could have had the body and the psychological characteristics which in the actual world are Swinburne’s, and Swinburne could have had the body and

psychological characteristics which in the actual world are mine' is to say that transworld identity can be what Garrett calls 'ungrounded'. It is to say, in other words, that given all the particular physical and psychological facts, these facts do not yet entail which person inhabits which body.

But the claim that the transworld identities of persons can be ungrounded is to be distinguished from, and does not entail, the claim that diachronic personal identity can be ungrounded. One might consistently claim that there is a non-actual (and therefore distinct) possible world just like the actual one except in that I have Swinburne's body and mind and he has mine, while *also* claiming that whoever has the body and mind which in the actual world is Swinburne's is such that, in *every* possible world in which that person has that body and mind at one time, that same person has that body and mind at every other time. To claim this would still be to maintain that DPI facts supervene on non-DPI facts; it would be to deny the Simple View, while affirming the thesis that there can be ungrounded transworld identities.

Therefore, if Swinburne (1984: 26) is right in, for example, asserting that 'the body which is presently yours ... could have been mine', this does not entail that the DPI facts are not fixed by the non-DPI facts; and so it does not entail the Simple View.

Nevertheless, it might be retorted that the argument for ungrounded diachronic identity still stands: for the challenge remains, for example, of justifying the assertion that the apparently coherent proposition that, after my death, I cross the memory-erasing Lethe, and am presented with a new body, really contravenes logic or the laws of nature. And even if *this* challenge can be fulfilled, a more conservative challenge would remain: to justify the assertion that, *contra* the two apparently plausible yet contradictory stories about who is *A* after Williams's operation (*vi*), at least one of these stories *cannot* be correct.

3.4.3 Responding to the broader argument (II): distinct imaginative presentations versus distinct possibilities

Let us now consider the case for thinking that, given a particular set of non-DPI facts, the proposition that *A* is the *A*-body-person, and the proposition that *A* is the *B*-body-person, represent *distinct* possibilities. The case hinges on being able to show that the first-person perspective enables us to see the distinction; for if all the non-DPI facts are the same, including the *A*-body-person's and the *B*-body-person's beliefs about

who they are, their tendencies to make certain statements and to behave in certain ways, *etc.*, then an *observer* certainly could not discriminate between the “two” possibilities; and Madell (1981: 99) seems to admit this.

The case for distinct possibilities seems to come to this. Suppose I am *A*, before the operation. Then I *can* imagine what it would be like for me to survive as the *A*-body-person; and I *can* also imagine what it would be like for me to survive as the *B*-body-person. And in imagining the former, I imagine something that is clearly *different* from what I imagine in imagining the latter. It seems to Madell, then, that there are two distinct possibilities regarding who is who, even though all the particular physical and psychological facts remain the same.

But it is necessary now to distinguish two kinds of difference: a difference in mode of imaginative representation; and a difference in state of affairs represented. Recall Subsection 1.6.2. We saw that in imagining a future situation, in which four people are playing pool, I might imagine this situation from various perspectives. I might imagine the situation from a bird’s eye view; or I might imagine it from the point of view of Benedict XVI (with whom, let us suppose, I shall be playing); or I might imagine it from my point of view. But what differs, in these various modes of imagining, seems to be this: that in imagining what the situation will be like for *x*, I imagine the very *same* situation, but in a way that privileges the perspective associated with *x*. Picturing the situation from *x*’s point of view, rather than from *y*’s point of view, certainly involves a difference in the *picture*; for example, it involves the fact that, in this picture, there is one person (*x*), whose head is never seen, unless in a mirror.

We may form such perspectival mental pictures of situations for various purposes. Suppose I am *A*. Consider the picture I form in *imagining what it will be like for me (after the operation)* if I am the *A*-body-person; and consider the picture I form in *imagining what it will be like for me (after the operation)* if I am the *B*-body-person. These are two distinct pictures, each of which I seem able to form. This seems to support Madell’s (100) claim that ‘*A* can easily form a conception of how his experience would be different’ in each story.

But we might form just the same pair of mental pictures for a different apparent purpose. You might ask me (*A*) to *imagine what it will be like for the A-body-person (after the operation)*; and you might then ask me to *imagine what it will be like for the B-person (after the operation)*. In imagining these things, I conjure

distinct mental *pictures*; but we do not say that on *that* account these pictures are of distinct *situations*.

Similarly, if tomorrow I shall play pool with Benedict XVI and Mel Gibson, I might imagine what our game will be like for Benedict; and I might imagine what our game will be like for Mel Gibson; and in so doing I conjure two distinct images. But it would be fallacious to infer, from the fact these images are distinct and able to be formed, that one represents truly one possible situation, but the other represents truly *another* possible situation. Rather, in this case, they are different pictures of the same situation.

In (3ZA) imagining what it would be like for me if I wake up tomorrow with Benedict's body and psychological features (in their totality, taking care to carry no memories over from my current mind), I generate exactly the *same* mental picture as I do when I embark on the exercise of (3ZB) imagining what it will be like *for Benedict* waking up tomorrow. But the distinctness of the mental picture generated in 3ZB from the mental picture generated in (3ZC) imagining what it will be like *for Harvey* (the actual me) waking up tomorrow certainly does not demonstrate that embarking on each of exercises 3ZB and 3ZC results in two mental pictures each of which corresponds to a distinct possible world. Therefore, given that the picture in 3ZA is phenomenologically indiscernible from that in 3ZB, any attempt to draw inferences from the distinctness of the picture in 3ZA from that in 3ZC, to the conclusion that these pictures represent distinct possibilities, looks highly suspect.

One may now wonder what possible situation *is* being represented by the picture I generate in imagining waking up with Benedict's body and mind tomorrow.

A reasonable line of thought is this. The picture generated in imagining what it will be like for Benedict waking up tomorrow, and the picture generated in imagining what it will be like for me were I to wake up in Benedict's body and mind tomorrow, are the same picture because they represent one and the same situation; but to describe this situation as one in which *I, Harvey* wake up in Benedict's body and mind is incorrect. For this is just the situation in which, undoubtedly, *Benedict* wakes up in his own body tomorrow.

The same picture, representing the same possible situation, might be described, on the one hand, as a picture of the situation in which Benedict, who is not me, wakes up in his own body (while I wake up in the same body I believe I have been waking up in all my life); or it might be described, on the other hand, but

erroneously, as a picture of the situation in which I wake up in Benedict's body and mind.

But now one might retort: 'But you are not Benedict; and so these are two *distinct* possibilities, are they not?'

Let us say that a person *x* is *Benedict-featured* if *x* has the body (incl. brain) which is today that of Benedict, and, furthermore, is fully psychologically continuous with Benedict as he is today. Let us say that *x* is *Harvey-featured* if *x* has the body (incl. brain) which is today Harvey's (mine), and is, furthermore, fully psychologically continuous with today's Harvey. Undoubtedly, the following two statements cannot both be true.

(3C) Tomorrow I shall be Benedict-featured.

(3D) Tomorrow I shall be Harvey-featured.

The following proposition is surely necessarily *false*.

(3CD) Tomorrow I shall be both Benedict-featured and Harvey-featured.

Why is 3CD necessarily false? Certainly, it is not the case that the second conjunct, 3D, is necessarily false. So either (a) there is no possible world in which tomorrow I shall be Benedict-featured, or (b) there are some possible worlds in which tomorrow I shall be Benedict-featured, but none of these worlds are worlds in which tomorrow I shall be Harvey-featured.

The important thing to note here is the availability of (a). It is extremely implausible to claim that, even in the absence of any physical tampering, the object which is now the computer chair I am sitting on could tomorrow become constituted by the very body of stuff that today constitutes the armchair in my loungeroom. Why should we be any less reluctant to rule out, as logically impossible, the analogous proposition with respect to persons? It seems that this lessening of reluctance arises from the fact that we can imagine various things from the first-person perspective. But the first-person perspective is just that: a perspective, from which a single possible world might be variously represented, not a distinguishing mark of possible worlds.

There are *two kinds of distinction* which survive a careful examination of Madell's (1981: 100) claim that there is 'a difference in experience of which *A* in the

first story can form a clear conception'. *Neither* of these kinds of distinction is the kind that Madell needs. But pointing out that there are these distinctions explains the temptation, which Madell seems to have fallen into, to think there is a distinction of another kind, *i.e.* between possible situations.

The first distinction is between pictures of situations. But two pictures of the same possible situation may differ; and so this is not the distinction Madell needs.

The second distinction is between two statements: '*A* is the *A*-body-person' and '*A* is the *B*-body-person'. These are distinct, and contradictory statements. They cannot both be true. But this does not entail that each is possible on its own. Similarly, the statements 3C and 3D cannot both be true. But it is plausible to think that 3C is the culprit, and that they cannot both be true because 3C cannot even be true on its own. I *cannot* be Benedict-featured tomorrow.

On these grounds, we may conclude that Madell's argument that *A* envisages two distinct possible situations, in envisaging now the situation that he will be the *A*-body-person, and now the situation in which he will be the *B*-body-person, is unconvincing.

We can make a similar response to Swinburne regarding the alleged possibility of personal persistence despite crossing the Lethe and being given a new body. Such stories seem consistent because we can form a sequence of first-personal pictures, some of which, the later ones, are radically different from others, the earlier ones. That we can string together such a sequence of imaginings enables us to follow such a story and to understand it. But we have just seen that the possibility of forming a certain first-personal picture does not entail that it is possible for *I*, the present imaginer, to be diachronically identical to the person privileged in the picture. Thus, though I can imagine what it would be like to have no memories (to occupy a body that has just crossed the Lethe), this does not entail that *I*, the imaginer, could be diachronically identical to a future person whose body has just crossed the Lethe. And if we think at any length about such a story, grave doubts arise about whether I could survive such a transmogrification. In what sense could the same person exist despite a total replacement of body and mind? The claim that it is possible for the same person to exist despite such changes seems, in fact, to be as implausible as the assertion that the swivelling chair I am now sitting on could persist though I destroyed its matter, and put in its place an armchair. There seems to be no way of making sense of my alleged survival despite such a radical change unless we *already* think that a further

entity – a Cartesian ego or substantial soul – is the real seat of personal identity, and that this substance persists despite all the physical and psychological changes. But we have good reason, as we have seen already, to *presume* the non-existence of such a further fact of DPI.

Recall, now, that Madell (1981: 100) thinks that even if *A* could *not* imagine how his experience would be different in the two stories, this would not entail that they are not distinct possibilities. And Swinburne claims that ‘only given verificationist dogma, is there any reason to suppose that the only things which are true are those of whose truth we can have evidence’. (1984: 24)

We may reply thus. If there is an alleged kind of thing, including a kind of fact, that we cannot empirically detect, we should want some motivation for positing it. We have noted this already, in Subsection 3.2.1. We need only add now that we have yet to be appropriately motivated.

We are tempted by the Simple View when we fail to distinguish variation between different imaginative presentations of a situation, from variation between different possible situations. It seems easier to confuse the two when considering first-personal representations of the world, than when considering impersonal representations. But the distinction is still there to be made.⁵⁹

3.4.4 Responding to the broader argument (III): distinct admissible precisifications versus distinct possibilities

Let us return now to Williams’s operation (*vi*). If there is a plausible story to tell, which tends to persuade us that *A* is the *A*-body-person, and another plausible story to tell, which tends to persuade us that *A* is the *B*-body-person, and *A* is not both the *A*-body-person and the *B*-body-person, then what explains our being persuaded, if it is not that *A*’s being the *A*-body-person, and *A*’s being the *B*-body-person, are distinct possible outcomes?

We have already seen one kind of answer to this question: that it is Williams’s rhetoric that persuades us: in ignoring the possibility of *A*’s and *B*’s conscientiously

⁵⁹ Johnston (1992: 592-594), too, traces belief in the Simple View to ‘the error of mistaking features of presentations of things for features of the things presented.’ (592)

believing in the physical criterion in Argument One; in glossing over significant differences in Argument Two; in underdescribing the cases in Argument Two; or in deploying soritical reasoning in Argument Two.

But there is another kind of answer to this question. And this second kind of answer is available even to one who, though s/he finds the Simple View unpalatable, remains genuinely puzzled by Williams's cases even when the second thought experiment is shorn of its (soritical or merely rhetorical) lead-up. For one might still think: 'But I would fear future torture were I told, nonquestionbeggingly, 'Your brain will be erased and then this body [pointing to the body which is presently mine] will have the following nasty things done to it.' And I am not convinced that this fear *would* go away were I given additionally the information about what will happen to *B*. But on the other hand, I also find the first story persuasive, and find it quite plausible to believe that body-swaps are possible.'

If we are puzzled in this way, our trouble would seem to be summed up as follows. Exactly one of the two following sentences is true, but we seem irremediably puzzled about which, and tend to waver, sometimes finding 3E almost irresistible, and at other times finding 3F almost irresistible.

(3E) *A* is identical to the *A*-body-person.

(3F) *A* is identical to the *B*-body-person.

Now, there are at least two possible explanations for our seemingly irremediable wavering. That we so waver is a good argument for the Simple View only if its explanation is the most plausible one. An indeterministic explanation is the other account.

The Simple Theorist's explanation says that we waver because we are ignorant about whether 3E or 3F is true, but one of these is determinately true. Given even a complete description of the non-DPI facts, we cannot infer either that 3E or that 3F is true, since the non-DPI facts do not entail the DPI facts. There are two (distinct) possible worlds in which these non-DPI facts are just as they are, one world in which 3E is true and another in which 3F is true; but we cannot, or do not, know which world is the actual one (*or*: which world would be the actual one if we performed the operation).

The indeterministic explanation says that there is only one possible world in which all the non-DPI facts are as they are. We are perplexed about which of 3E and 3F is true because the meanings of our expressions, in particular of ‘person’ and/or ‘is the same person’ are very vague, in that *both* a physical criterion *and* a psychological criterion is associated with these terms, and our language is not so precise as to determine that one of these criteria is uniquely correct. This is unsurprising given that we have developed the meaning of ‘person’ and ‘is the same person’ in a world devoid of information-swaps between brains. We have never had to attach to the predicate ‘is the same person’ any rules for adjudicating who is the same person as who in such hypothetical cases as Williams’s. So no rule is available for us to consult in struggling with the question of how to respond to such cases. Our language, and the non-linguistic facts, leave some statements indeterminate, as we saw in Chapter 2; and we might reasonably argue that one of the statements which they leave indeterminate is the conclusion of Argument One (the contradictory of which is the conclusion of Argument Two). We might reasonably maintain that *A* is identical *either* to the *A*-body-person, *or* to the *B*-body-person (and that *B* is identical to the person with the body that remains), but that it is indeterminate *which* he is identical to.

So, according to the indeterministic explanation, there is only one possible world in which all the non-DPI facts are as they are; but on one admissible precisification of our language, this world is one in which 3E is satisfied and 3F is not; and on another admissible precisification 3F is satisfied and 3E is not. Since on any admissible precisification, the DPI fact is entailed by the non-DPI facts, it is true *simpliciter* that the DPI fact is entailed by the non-DPI facts.

The indeterministic explanation of our puzzlement is more plausible than the Simple Theorist’s explanation because it is more ontologically parsimonious. The Simple Theorist’s explanation posits a further kind of fact, and one for which we have no empirical evidence. Furthermore, no *a priori* argument in its favour so far examined has stood up to scrutiny.

Thus, even if Noonan’s, Garrett’s, Beck’s and others’ critiques of Williams’s Arguments One and Two, and even *if* the above critique of the broader argument for the Simple View, leave one still genuinely puzzled about whether to believe 3E or 3F, this is *still* not sufficient reason to believe the Simple View.

The claim that it is genuinely indeterminate which criterion of personal identity (physical or psychological) is correct is explicitly defended by Sider (2001),

who takes pains to argue that certain conditions must be met before a philosophical conundrum can reasonably be resolved in this way, and that the particular conundrum regarding physical *versus* psychological criteria of personal identity satisfies these conditions. I find Sider's argument fairly persuasive, but I will not reiterate it or defend it here. What I think I have established here is just that, in the event that Argument One and Argument Two are, all things considered, literally just as good as one another, Sider's response would be a *less* implausible response to this outcome than would the Simple View.

Having said this, we should note that this indeterministic approach to operation (vi) would seem to commit one to a claim of the following form: 'Either x at time t is identical with y at distinct time t^* , or x at time t is identical with z at t^* , but it is indeterminate whether x is y or x is z .' This is also the form of the particular indeterministic response to the *Fission Case* that I defend, in Chapter 4, from criticisms by Garrett (1998). Some of these criticisms might also be thought applicable to the claim that operation (vi) gives rise to indeterminate SDPIs, in virtue of this common form. Having defended the indeterministic view of the Fission Case from these criticisms, however, it should be clear that an indeterministic view of operation (vi) could be defended along similar lines.

Regarding operation (vi) I claim only that an indeterministic view is more plausible than the Simple View. But I shall argue in Chapter 4 that, in the case of Fission, an indeterministic response is more plausible than *any* other response.

3.5 OTHER RESPONSES TO WILLIAMS

For the sake of completeness, other responses to Williams's conundrum deserve brief mention here, though they are tangential to the main task.

One response is to argue that the contradictory nature of our reactions to such science-fiction cases shows that our intuitions are unreliable when confronted with such cases, and that the very methodology of consulting thought experiments is "X" – where 'X' might be replaced by anything ranging from 'bankrupt' to 'to be employed only with great care'. The more extreme version of this line is defended by *e.g.* White

(1989), who tends toward the “bankruptcy” allegation; while Johnston (1987, *e.g.* p. 81) seems to defend a more moderate, qualified version of this line.

Beck (1998 §5) argues that White’s response is too extreme. The methodology of such cases would be suspect *if* their goal was ‘to elicit from us philosophically correct intuitions’, for sometimes different intuitive responses cannot all be correct. But a more nuanced understanding of such thought experiments is that they elicit, through our intuitive responses, the criteria implicit in our concepts. These criteria may sometimes conflict; but that we are pulled to some extent in *each* direction is consistent with our intuitive response in one particular direction being the *dominant* response. In cases such as Williams’s, ‘implicit criteria which are usually co-satisfied can be made to come apart, with the possible result that we discover to which underlying principles we are more strongly attached.’ (Beck 1998 §5) Beck thinks our dominant response to Williams’s twin thought experiments is to think *A* is the *B*-body-person, showing, presumably, that we are ‘more strongly attached’ to a psychological criterion of DPI.

Beck seems to place great significance on the fact that some one response to Williams’s operation (*vi*) is dominant. He writes: ‘the existence of a case which shows a generally felt, direct inversion of intuitions like that under discussion would be a serious threat to the usefulness and reliability of the method, since it raises the suspicion that should any given thought-experiment be differently described we might feel an intuition totally opposed to the one we currently feel.’ (1998 §5) Beck goes on (in §§6-7) to ‘avert the threat Williams’s examples pose to the coherence of the concept of personal identity in particular’ (§5) by arguing, as we have already seen, that the mind-swap intuition is elicited by question-begging descriptions and other faulty aspects of Argument Two (*e.g.* glossing over large differences in cases, or soriticality).

However, Beck overstates the threat. Even if our intuition that operation (*vi*) constitutes a mind-swap remained strong in the absence of the faulty Argument Two (and for some it has seemed so to remain), and even if at the same time the intuition that (*vi*) is a body swap at other times seemed to pull us *just as strongly* in the opposite direction, this would be fatal neither (a) to the methodology of thought experiments, nor (b) to the coherence of the concept of personal identity.

Regarding (a), very many thought experiments do not generate multiple contradictory, equally forceful intuitions, even if they generate seemingly

interminable debate between contradictory *considered* responses. Other thought experiments generate broad consensus in both intuitive and theoretical respects. Uncontroversial thought experiments tend not to attract as much attention as controversial ones, and so we are perhaps under the impression, at times, that irresolvable conflict is more pervasive than it actually is.

But even when both intuitive reactions and considered responses to a particular thought experiment seem both irreconcilable and equally forceful, such a conundrum still need not render that thought experiment philosophically useless. It may rather shed light on the multiplicity of different criteria attached to the concept under investigation.

Moreover, if there are multiple equally weighty criteria attaching to a particular concept, which are in conflict in a particular case, this does not automatically render that concept incoherent. For regarding (b), Williams's case does not threaten the 'coherence of the concept of personal identity', even we are pulled equally strongly in each of the conflicting directions. Firstly, the criteria do not conflict in everyday reidentifications of persons. Secondly, in operation (vi), assuming even that the criteria pull on us each with equal force, it is *not* the case that we intuitively think, '*A* is *both* the *A*-body-person and the *B*-body-person.' *That* response would be *prima facie* grounds for suspecting our concept of personal identity of incoherence. (And even here, a contradiction could be acceptable on a paraconsistent understanding.) But that is *not* our response. Our response is: '*A* is the *A*-body-person or the *B*-body-person, *but not both*' – even if we seem irremediably uncertain about which he is. And since he certainly is not both, our concept of personal identity is certainly not incoherent, even if it is very vague, in the way discussed in Subsection 3.4.4.

3.6 CONCLUSION

The Simple View is implausible. In the absence of a good argument in its favour we should reject it. Williams's (1970) twin thought experiments are not the basis for any such argument. Neither is the "broader argument" persuasive.

The Simple View seems plausible only when we confuse multiplicities of acceptable imaginative or linguistic representations of a single possible world with

multiplicities of possible worlds. Failure to recognise such merely presentational variability in relation to questions of diachronic personal identity leads one to infer a multiplicity of worlds, as the only explanation for one's inability confidently either to affirm or to deny certain SDPIs, even when one knows all the relevant particular physical and psychological facts.

In Subsection 4.3.6, we shall consider briefly whether the so-called *Fission Case* affords a persuasive argument for the Simple View, as has been suggested by Swinburne (1984: 13-20). We shall see that it does not.⁶⁰

3.7 SOME FURTHER REMARKS

The investigations of the past two chapters suggest that the relationship between statements of diachronic personal identity, linguistic conventions, and the extra-linguistic world, has the following features.

Given an SDPI $\langle x, y, t, t^* \rangle$, once all the particular physical and psychological facts have been fixed, there is left no room for variation regarding *whether* $x=y$. The only room for variability that may be left concerns the *symbol* ' $x=y$ '. That is, where ' $x=y$ ' is indeterminate, there may be variation among different admissible precisifications of 'person' (or 'same person'), such that on some such precisifications, ' $x=y$ ' is true, and on others, ' $x=y$ ' is false.

Suppose ' $x=y$ ' is indeterminate. Then we might *settle* the meaning of 'same person', so that ' $x=y$ ' is no longer indeterminate. (See Subsection 2.4.4.) We would then have effected a refinement in the meaning of the symbol ' $x=y$ '. (We would have added to the application rules that are widely assumed not to be followed: see Subsections 2.7.3 and 2.7.6.) But this would not effect a change regarding *whether* $x=y$. *Whether* $x=y$ depends on particular physical and psychological facts; but it does not depend on language. Thus, it would be incorrect to say that the question of who is actually who depends on linguistic conventions.⁶¹ In contrast, the question of whether,

⁶⁰ For an interesting attack on the Simple View that is based on variations on the Combined Spectrum Argument, see Shoemaker (2002).

⁶¹ See Merricks (2001) for an argument that 'personal identity over time is never a matter of convention.' (173) (I do not mean to endorse Merricks's argument.)

at any particular point in time, the *symbol* ' $x=y$ ' is an indeterminate SDPI, clearly *does* depend on the linguistic conventions in force at that time.

The important point here is this. Once the new conventions are settled, we are operating within those conventions: these are now the conventions in virtue of which our utterances mean what they do. So we cannot, after the settling, look back and utter truly this sentence: 'I would not have been so-and-so had our language not been settled in this way.' We should rather say: 'Had our language not been settled this way, it would have been wrong for me to say 'I am so-and-so'.'

The thought that Personal Indeterminism commits one to thinking that who *is* who depends upon linguistic conventions has seemed to cause some to be uncomfortable with Personal Indeterminism.⁶² But this thought is seen to be mistaken once we draw carefully the distinction between (a) variation concerning different factual possibilities, and (b) the *symbolic* variation concerning different ways a single set of facts might admissibly be described or presented – a distinction that we have seen to be crucial in enabling both the epistemic view and the Simple View to be resisted.

Consider that whether I *am* bald does not depend upon how 'bald' is used, but only upon the physical facts about my scalp. (See Subsection 2.7.2 on the distinction between entailing and determining, in relation to the example of 'bald'.) Supposing I dread becoming bald, it would make no sense for me to be relieved were people to change the use of the word 'bald' so that the sentence 'You will probably become bald, like your father and your grandfather' now counts as false, instead of true. Yet this is not usually seen as a reason to reject the claim that there can be indeterminate predications of 'bald'. For although the conventions for describing the relevant facts have changed, these facts have not themselves changed: my scalp remains just as likely to become bald as it always was; but it is now less likely to be called 'bald'. Compare Williams's (1970: 178) observation regarding the apparent inappropriateness of relief simply in virtue of a refinement of the meaning of 'same person' so that the sentence 'You will be tortured' now counts as determinately false, instead of indeterminate. (See Subsection 1.6.6.) It is clear now why this observation gives us no reason to reject Personal Indeterminism.

⁶² See Williams (1970: 178). Nozick (1981: 96-97) seems to have had similar thoughts, as does Merricks (2001).

So far I have used two locutions synonymously: (i) ‘‘*S* is indeterminate’; and (ii) ‘it is indeterminate whether *S*’. Given that indeterminacy is essentially a property of *symbols*, not of states of affairs, (i) should be regarded as the basic locution, of which (ii) is merely a derivative form.

Chapter 4

FISSION

4.1 INTRODUCTION

I say that an *indeterministic response* to a given hypothetical operation performed on a person or persons is one which alleges that that operation gives rise to indeterminate SDPIs. We have so far examined two hypothetical operations that might be thought to give rise to indeterminate SDPIs: an arbitrary central operation in the Combined Spectrum; and Williams's (1970) operation (*vi*). I have argued that the former certainly does give rise to indeterminate SDPIs; and I have argued just that it is *more* plausible to think the latter gives rise to indeterminate SDPIs than to think it supports the Simple View.

In the present chapter, however, I will argue that it is *most* plausible to adopt an indeterministic response to another hypothetical operation: that of the *Fission Case*.

The Fission Case, at least in its modern guise, was first described by Wiggins (1967), although it had a precursor in Williams's (1956-7) "Guy Fawkes Case". The case has been significant for two reasons. Firstly, it has been seen as a testing ground for various theories regarding the necessary and sufficient conditions for personal identity over time. As such the pertinent question concerns *who is who* in the Fission Case. Secondly, some (notably Parfit 1984 esp. Ch. 12) have thought that consideration of the Fission Case sheds light upon certain normative matters, especially the question of whether the personal identity relation, or some other relation(s), is of greater normative significance.

I shall be interested principally in the question of who is who in the case, although I will touch upon the normative question in places.

In Section 4.2, I shall describe the Fission Case, and will draw some initial conclusions concerning a related concept, that of a brain transplant. In Section 4.3, I will survey and critique alternative accounts of who is who in the Fission Case, and will give reasons for preferring a particular indeterministic response to the case. In Section 4.4, I shall respond to four criticisms by Garrett (1998: 64-67) of the indeterministic response. In Section 4.5, I will consider the nature of the debate concerning who is who. I will argue that practical considerations have some relevance to this debate, and that intuitive normative claims concerning *who deserves what* in the Fission Case seem to complement the indeterministic response to fission.

4.2 THE FISSION CASE

4.2.1 Preliminaries

Some preliminaries are in order before describing the Fission Case proper. Very plausibly, the following are at least *possible* states of affairs, violating neither logic nor the laws of nature.

(4A) A body may be kept “alive”, merely in the sense that its heart keeps beating, its lungs keep breathing, *etc.*, though its brain has been removed.

(4B) Just as modern surgery enables heart transplantation, some possible surgery of the future enables brain transplantation. A brain may be transplanted, that is, into the skull of an awaiting “living”, brainless body, and connected to that body; and that body and brain will then belong to a fully functioning human being the brain of whom communicates with the limbs *etc.* as per usual. Further, given the continued viability of the transplanted brain, given that it is only *moved*, but not internally tampered with, the human being who has the transplanted brain after the transplant is fully psychologically continuous with the human being who had that brain before the transplant.

(4C) For some person x , x 's whole brain contains such functional redundancies and informational duplication that, for at least one of its hemispheres, were this hemisphere excised and destroyed, there would still be virtually as much psychological continuity as if the hemisphere had remained intact. Psychologically speaking, it would be virtually as if x 's brain had not been tampered with at all. The removed matter might be replaced with some inert substance of equivalent weight, so

that x 's head does not feel lopsided or strangely light to x ; in fact, a surgeon might anaesthetise x in x 's sleep and perform such an operation so that when x awakes x is not aware that there has been any surgery at all.

(4D) Some people's brains are symmetrical, each hemisphere having similar capabilities. (Noonan 1989: 5; Garrett 2004 §3) (4D*) For some persons, either hemisphere could sustain psychological continuity on its own.

(4E) For some (otherwise normal) individual human being x , the transplantation of *either* hemisphere of x 's brain into an awaiting brainless living body b would be such that, after the transplant, a person would occupy b which person would be psychologically continuous with x (or which person's post-transplant stage would be psychologically continuous with his/her pre-transplant stage, or (some other locution appropriate to some plausible view about what the relation of the psychological continuity relation are)).

4.2.2 Fission

Given 4A-4E, the following scenario is surely possible, contravening neither logic nor the laws of nature. I am one of three brothers who are monozygotic ("identical") triplets. My brothers are called 'L' and 'R', and I am called 'V'. A short time ago, each of my brothers' brains was been wholly removed from his skull and destroyed, but their bodies have been kept "alive". Let us call these bodies the *L-body* and the *R-body*, their skulls the *L-skull* and the *R-skull*. My brain is symmetrical. An event now occurs which may be called 'my fissioning'. The left hemisphere of my brain is transplanted into (and connected up to) the L-body; and the right hemisphere of my brain is transplanted into the R-body. The hollow half of the L-skull is filled with some inert material of the same density as a brain, as is the hollow half of the R-skull. My pre-fission body is destroyed. This is the Fission Case.⁶³

After I have fissioned, two conscious entities with separate mental lives have, respectively, the L-body and the R-body. They might move to different countries and carry on different lives. But, after the operation, each exhibits psychological continuity with me, V, as I was before the operation. We may call the entity with the L-body 'Lefty', and the one with the R-body 'Righty'.

⁶³ The Fission Case is due to Wiggins (1967), although there are minor differences between my description and Wiggins's.

4.2.3 Some initial conclusions

The following claims seem to be agreed upon by *almost* every participant in the modern debate surrounding the Fission Case.

Consider the case in which my entire brain is moved to a new body *b* and connected up to it, as under 4B. Am I now the person whose body is *b*? It is hard to deny that I am. Parfit is surely right in saying that '[r]eceiving a new skull and a new body is just the limiting case of receiving a new heart, new lungs, and so on.' (1984: 253) This strong intuition arises from what is surely the central role of the brain in supporting, in ordinary lives, that which is so central to the notion of personhood: a continuing mental life.

What is important about my brain, in respect of my own survival, is that it supports my mental functioning. Suppose my brain is symmetrical as under 4D. Then my mental functioning would not be significantly diminished by the destruction of either hemisphere of my brain. Surely *I* should then persist through the destruction of either hemisphere.

Given all of the above, it is very hard to deny that were either hemisphere of my brain destroyed, and the remaining hemisphere transplanted into an awaiting de-brained living body *b* (and connected to it), I should be identical to the person whose body, after the operation, is *b*. Such an operation would be an *effective* brain transplant with the same outcome for questions of my diachronic identity as the transplant of the whole of my brain.

In general, then, I should persist through a procedure in which my entire brain is transplanted, or in which half of my brain is destroyed and the other half transplanted. I should certainly not exist in my old body, which we might in any case suppose to have been destroyed. I exist, rather, with a new body. I have a living, functioning brain and body both before and after the procedure, and psychological continuity is sustained by the continuing existence of a certain portion of brain, and so it is surely correct to say that I *survive* the procedure. We can summarise this conclusion thus: Effective brain transplants are in general not fatal, and may in general more properly be described as body transplants, since in general I go where my brain goes.

4.3 RESPONSES TO THE CASE

4.3.1

Fission has been seen as a test case for various theories regarding the necessary and sufficient conditions for diachronic personal identity. As such, it invites the question *who is who* in the case.

Garrett (1998: 59-67) distinguishes six answers to this question, one of which is the indeterministic response. None of these answers is *obviously* correct. In Subsections 4.3.2–4.3.6, I shall describe the five answers that I reject, giving reasons for disliking each of them. In Subsection 4.3.7, I shall outline the initial case for favouring an indeterministic response to Fission.

4.3.2 Claiming that the fission of persons is impossible

If personal fission were impossible, there would be no need really to confront the question of who would be who were the case carried out.

However, as Garrett (1998: 59-60) points out, the view that fission is impossible is very implausible. Fission seems merely technically impossible. Given that we can transplant organs and limbs, there seems to be no barrier in the laws of *nature* that would prevent our transplanting a whole body (*i.e.* transplanting a whole brain). And there seems to be nothing logically incoherent in the description of the particular physical events that would constitute the transplantation of a whole brain. Given that actual people do have symmetrical brains, there seems to be no deep barrier to the possibility of splitting a brain and transplanting each of its halves into a separate de-brained body, so that each hemisphere sustains psychological continuity with the original whole.

One might argue that the splitting of the *lower* brain is impossible. But as I pointed out in Subsection 1.4.9, no such impossibility would seem to arise from the very laws of nature. The apparent impossibility of dividing the lower brain (while preserving sufficient functioning in each half) would seem to be of a shallow sort, a product of the particular design of our brains, in combination with the particular respects in which surgery is technically limited. It would not seem to be ruled out by logic or the very laws of nature.

We should reject the claim that the Fission Case is impossible.

4.3.3 Claiming that V survives as both Lefty and Righty

The most straightforward way of claiming that V survives as both Lefty and Righty is to claim that V is *identical* both to Lefty and to Righty. But this commits one to the thesis that Lefty is identical to Righty, on account of the transitivity of numerical identity. But Lefty and Righty have separate bodies and mental lives, making this claim very implausible. Though initially quite similar to one another, they will, as Garrett points out ‘soon begin to differ, mentally and physically’. (1998: 61) They could move to opposite sides of the world, have separate careers, *etc.*. So we should reject the view that V is *identical* both to Lefty and to Righty.

Perhaps one might claim that Lefty \neq Righty, while claiming that V survives “as” both Lefty and Righty in the sense that each of Lefty and Righty is a *component* of V. One might claim this either: (a) by claiming that Lefty and Righty are not persons, but are parts of the single person V; or (b) by claiming that Lefty and Righty are distinct persons, but together constitute a third person, V.

But as Garrett (1998: 61) points out, (a) and (b) are both very implausible.

Each of Lefty and Righty looks, acts, thinks and feels just like a whole person. In other words, ‘Lefty and Righty both satisfy the normal physical and psychological criteria for personhood.’ (Garrett 1998: 61) So we should reject (a), which denies that Lefty and Righty are persons.

Claim (b) is also implausible. Once we have acknowledged that each of Lefty and Righty is a distinct person, the claim that they together constitute a *further* person is rather difficult to believe. There are only two separate minds, and only two bodies. It is very difficult to see where a third person fits into the picture. The only motivation for positing the additional person would be a belief that no other response to the Fission Case is plausible. But we shall see that such a belief would be false.

We should reject the claim that V survives as both Lefty and Righty.⁶⁴

4.3.4 Cohabitationism

Some have thought that, in fact, Lefty and Righty, though distinct persons, *both* exist prior to fission, though in exactly the same space, sharing the same body (and

⁶⁴ In Subsection 4.3.5 I will consider a loose sense in which V might be thought to “survive as” both Lefty and Righty while literally going out of existence at fission.

presumably the same mind). Fission causes these persons to become spatially distinct. (See *e.g.* Lewis 1983, Noonan 1989.) This view may be called ‘cohabitationism’.

Cohabitationism is implausible.

Firstly, it could be that I myself will one day fission. On the cohabitationist view, if I will one day fission, then there are two⁶⁵ persons inhabiting my body right now. But if I will never fission, then there is only one person inhabiting my body now. It seems odd that the number of persons here now should depend on whether such a future event occurs. (Garrett 1998: 63)

Secondly, to say that two persons could be in exactly the same place at exactly the same time, and share just the same body and mind, seems to be, as Garrett puts it, ‘a tremendous distortion of our concept of a person.’ (1998: 62)

Suppose I am (or seem to be) the only person now in this room, but that one day I will fission. Some cohabitationists (Lewis 1983: 152-155, Noonan 1989: 141-142) have argued that even if, strictly speaking, two persons now share my body and are now in this room, it may still be correct to answer ‘One’ to the question ‘How many persons are now in this room?’ They have proposed an alternative “method of counting” according to which this answer is correct. This method allows that, if we are counting the number of persons *at a given time*, we count not the number of persons *per se*, but the number of *person-stages* existent at that time. The notion of a person-stage is a *perdurantist* one. (See Subsection 1.7.5.) A person is regarded by perdurantists as an aggregate of person-stages (satisfying certain conditions). For each time, I have a person-stage which exists only at that time. Although two persons (strictly speaking) are in this room, they now share the same person-stage. Counting by person-stages, rather than by persons, ‘There is one person in this room’ can be regarded, so it is argued, as a correct statement, since there is only one person-stage now in this room.

This proposal does not really reduce the counterintuitiveness of cohabitationism. For according to this doctrine, there *are* two persons in this room now, sharing the same body and mind; and *that* is the counterintuitive claim, to which cohabitationists, by definition, are committed. If we re-define ‘person’ as ‘person-stage’, and then assert ‘There is only one person now in this room,’ we have

⁶⁵ Some cohabitationists, *e.g.* Noonan (1989), think that *three* persons share V’s body prior to fission, one of whom terminates at fission.

distracted attention from the counterintuitive claim, perhaps; but we have not made *it* any less counterintuitive.

Lewis (1983: 153) argues that, in certain contexts, we do count by stages or segments, though our form of words superficially suggests we are counting by wholes. For example, he says, a certain stretch of highway is part of both Chester A. Arthur Parkway and Route 137. Yet to cross that stretch is, we say, to cross *one* road, not to cross two.

We might argue that Route 137 is not in fact a road, but is rather a *route* – a slightly different concept. We could then argue that, in this case, we are counting not by road-segments but by roads, thus defeating the force of Lewis's example.

However, we need not argue this way. Even if Lewis's example is really one of counting by segments rather than by wholes, the fact remains that the claim that two *persons*, strictly speaking, can share one body and mind, is quite difficult to swallow. In contrast, the claim that two roads can share one segment is not really counterintuitive.

Though there is no watertight argument against the cohabitationist response to fission, it is quite an odd view, and we should reject it if there is a more intuitive alternative response. And we shall see that there is one.

Incidentally, a cohabitationist who, like Lewis (1983), thinks that just *two* persons coexist before fission seems committed to the view that a statement such as 'I am identical with Lefty' is indeterminate, if uttered by V, since 'I' would seem, for such a cohabitationist, to be indeterminate in denotation, ranging over Lefty and Righty. (See Noonan 1989: 230.) This seems the only way for the cohabitationist to make sense of the very plausible claim that such a statement, uttered by V, would be meaningful and significant. Such a cohabitationist would seem committed, then, to an indeterministic response to the Fission Case.⁶⁶ However, my argument for an indeterministic response does not at all depend on this claim. Note, also, that the reverse claim does not hold.

⁶⁶ Appealing to indeterminacy may be useful, too, for a perdurantist who believes, with Lewis (1983), that in the case of the very long-lived Methuselah, an infinite number of overlapping persons cohabit the Methuselah-body at a particular point in time. The question, 'Who am I?', uttered by the Methuselah-mouth, may be answered thus: 'Determinately, you are one of the many present cohabiters of the Methuselah-body, but it is indeterminate *which* one you are.' This answer affords a partial reply to a critique by Campbell (2001) of Lewis's stance on Methuselah, one aspect of which (see 2001 §6) is the allegation that the aforementioned question cannot satisfactorily be answered.

For the indeterministic response I shall be defending does not commit me to the view that more than one *person* shares the V-body: see Subsection 4.4.3.

4.3.5 Claiming that V ceases to exist at fission

Some have argued that V is neither Lefty nor Righty, and that V ceases to exist at the point of fission. (See *e.g.* Nozick 1981, Shoemaker 1984 Ch. 12, Garrett 1998 Ch. 4.) Garrett sums up the motivation for this view as follows.

When I divide, there are two equally good candidates for identity with me. Since they are equally good, and since one thing cannot be two things, I am identical to neither. And since there is no one else with whom I could plausibly be identified, I no longer exist after fission. This response respects the logic of identity, and does not violate our concept of a person by supposing either that two persons compose one large, scattered post-fission person or that more than one person occupies the pre-fission body. (1998: 63)

This view, which may be called the *orthodox best candidate theory*, seems a more natural response to the Fission Case than any of the others so far examined. In particular, on this view, there is, at any point in time, just one person for each human body and mind. So this view does not distort our *synchronic* concept of a person (*i.e.* our concept of a person at a single point in time).

However, the orthodox best candidate theory is committed to quite a counterintuitive claim, *viz.* the claim that fission terminates V. To see the oddness of this claim, consider that, had only one hemisphere been transplanted, and the other destroyed, V would determinately have survived: effective brain transplants, as we have seen in Subsection 4.2.3, are not generally fatal. But if an ordinary brain transplant is not fatal, why should fission be fatal? Fission is a brain transplant that is doubly successful. The sort of event that constitutes the death of a person is generally the sort of event that causes the cessation of certain processes, and in particular of the processes supported by brain matter. But no such event occurs in fission.

One might argue that there is a *sense* in which V “survives” fission. (See Parfit 1971b.) One might argue that Lefty and Righty, since they are psychologically continuous with V, are related to V in all those ways that *matter* to V. Since Lefty and Righty could carry out V’s wishes, could cherish memories (or apparent memories)

inherited from V, could ensure that V's unfulfilled projects are completed, *etc.*, one might argue that V's fissioning is *as good for V as* his surviving normally. But to claim that V *literally* survives fission – that fission does not kill him – while also claiming that he is caused by fission to go out of existence, seems to be to distort the meaning of the word 'survive'. For a non-existent person surely cannot *be* alive except in a metaphorical sense; and one cannot have literally survived an event that has caused one to cease living.

The motivation behind the view that fission kills V, as expressed in the passage quoted from Garrett above, is the desire to respect the transitivity of the identity relation, without violating our concept of a person. But in trying to satisfy this desire, orthodox best candidate theory abandons what seems a quite intuitive component of our *diachronic* concept of a person. That is, (4XX) we ordinarily would think that, if enough of a person's brain continues to exist and to support a continued mental life, that person will continue to exist. This intuitive thought underlies the beliefs seen in Section 4.2 to be so plausible: that brain transplants are generally not fatal; that a person generally goes where his/her functioning brain goes; that a person with an appropriately symmetrical brain can survive the destruction of either hemisphere; and that such a person can survive the destruction of one hemisphere and the transplantation of the other. We should not abandon the intuitive claim 4XX unless we are forced to do so.

To bring out the plausibility of 4XX, consider what one's attitude would be if one were faced with the prospect of fission. Would one regard this as impending death?

It would seem odd to *fear pain* that occurs after one has ceased to exist. But I think that, if I faced the prospect of fission, and I knew that *both* fission products would be tortured after the operation, I would not feel any less fearful of impending pain *at all* on account of the prospective existence of two arenas of tortured consciousness, rather than of one. I do not think I could rightly be accused of irrationality in feeling this fear. But my fear seems inseparable from the feeling that I will, in some very real sense, still be around to experience the pain. My fear here is of the experience of pain, not of impending death.

Even if, as Garrett (1998: 92-93) suggests⁶⁷, one may have good reasons to prefer ordinary (non-branching) survival over fission, this does not entail that it is *any* less reasonable to be especially concerned about, and fearful of the pain to be experienced by, one's fission products, given that they will both exist, than of the pain to be experienced by one's ordinary future self. Concern regarding (non)existence is not the same as concern regarding the quality of experience.

One might argue, as *e.g.* Parfit (1971b, 1984 Chs. 12-14) and Shoemaker (1970, 1984 Ch. 13) have done, that such considerations support the claim that some relation other than personal identity is what really underpins rational special concern for particular future selves: one might maintain that V ceases to exist, but that he is still rational in fearing (or being especially concerned about) the impending pain, since he is linked to the torturees by relations of psychological continuity and/or connectedness, which, unlike the DPI relation, is/are normatively significant. But that would be a surprising conclusion. It is more natural to think that V is rational to fear impending pain because he will still be around to experience that pain. If the logic of identity, and our concept of personhood, will accommodate this thought (and we shall see they can), the surprising conclusion is unmotivated.⁶⁸

Finally, as Rieber (1998: 584) points out, 'division does not seem to be cessation of existence.' In general, when we see something dividing, symmetrically or asymmetrically, into two – be it a river, the branch of a tree, or a puddle drying up gradually, on uneven ground, until two shallow pools lie where once one deep pool stood – we seem to think and to speak of it as having *split into*, and in some sense *become, two*. This suggests that to divide is to *undergo* a process.

If some philosophical account of personal fission can validate and make sense of this pretheoretic way of construing division, then we ought, other things being equal, to prefer that account. The orthodox best candidate theory, however, invalidates, rather than makes sense of, the notion that, in fissioning, a thing undergoes a process, by which it in some sense becomes two. On the orthodox best

⁶⁷ Garrett makes this point in the course of critiquing Parfit's (1984 Ch. 12) argument from fission to the claim that personal identity "does not matter".

⁶⁸ Johnston (1992) uses an indeterministic view of fission as the basis for an attack on Parfit's (1984 Ch. 12) argument from the Fission Case to the claim that personal identity "does not matter". However, Johnston's argument here is different to mine. He argues that the special concern one ordinarily has for oneself can be "extended", in the case of one's fissioning, to one's prospective fission products. He does *not* claim, as I do, that: V determinately survives fission, and may thus literally be said to care for *himself*, in caring jointly for Lefty and Righty.

candidate theory, a thing that splits symmetrically does not undergo or become anything; it simply goes out of existence. To cease to exist is not to *undergo* a process: to undergo a process, a thing must exist both at the beginning and at the end of that process.

None of the above considerations weigh conclusively against the orthodox best candidate theory; but they do show that that theory has its costs. We should prefer a view that avoids the costs, while still satisfying the motivating desire, *viz.* to respect the transitivity of identity without distorting our concept of a person.

4.3.6 Claiming that either V is determinately Lefty or V is determinately Righty

If all the preceding options are rejected, only one option seems to remain: to claim that either V is Lefty, or V is Righty.

This claim has seemed to many to be implausible, on account of the symmetry of the Fission Case in all particular physical and psychological respects. One wonders what could possibly *make it that* V is, say, Lefty but not Righty.

If one holds the Complex View, it is very hard to make sense of the claim that any fact – any aspect of the world – could *determine* that, say, V is Lefty rather than Righty. For on the Complex View, questions of personal identity depend wholly on the particular physical and psychological facts; and all these facts are symmetrical in the Fission Case, making neither Lefty nor Righty the sole best candidate for identity with V.

If one rejects all the other responses to the Fission Case, and is led to conclude that it *must* be the case that V is either Lefty or Righty, then one might, given the above reasoning, think, as Swinburne (1984: 13-20) does, that the Fission Case motivates the Simple View. For if the Simple View is true, then, even though the particular physical and psychological facts are all symmetrical, there is a *further* fact of DPI, and this fact is not mentioned in the symmetrical description of the Fission Case. The existence of this further fact, then, would explain what could *determine* that V is, say, Lefty but not Righty.

However, we have seen in Chapter 3 that the Simple View is implausible and ought to be rejected unless there is a good argument in its favour.

Suppose we agree that V is either Lefty or Righty. Is this proposition the foundation of a good argument for the Simple View? We have seen that ‘(determinately) either p or q ’ does not generally entail ‘either determinately p or determinately q ’: the determinacy operator is “non-constructive”. (See Subsection 2.5.1.) Therefore, the claim that either V is Lefty or V is Righty does not entail the claim that either V is determinately Lefty or V is determinately Righty. We can consistently claim that V is either Lefty or Righty, while denying that the facts *determine* that V is Lefty, and denying that the facts *determine* that V is Righty. Thus we can deny that there is a further fact, a further aspect of the world, which, despite the symmetry of the particular physical and psychological facts in fission, is asymmetrical in regards to Lefty and Righty.

Hence, the claim that V is either Lefty or Righty does not ground a good argument for the Simple View, since it does not entail the existence of a further fact of DPI. This conclusion is conditional, however, upon the coherence and plausibility of an indeterministic response to the Fission Case. For given the claim that V is either Lefty or Righty, denying both that it is determinate that V is Lefty and that it is determinate that V is Righty commits us to the claim that it is indeterminate whether V is Lefty and indeterminate whether V is Righty. (We shall see in more detail why this is so in Subsection 4.3.7.) It thus commits us to an indeterministic response to the case. If one disagrees with the Simple View, but is not prepared to accept an indeterministic response, then, one must deny that V is either Lefty or Righty. In Section 4.4 we shall consider whether the indeterministic response can withstand certain criticisms by Garrett (1998). Presently, we shall examine the positive case for an indeterministic response.

4.3.7 The indeterministic response

Claiming that V (determinately) survives fission, though it is indeterminate whether he is Lefty or Righty, is one kind of indeterministic response to the Fission Case. There is another kind of indeterministic response though: claiming that it is indeterminate whether V survives fission. I see no motivation for the latter claim, however, and shall ignore it henceforth.

I claim just that it is determinate that V survives fission, but that it is indeterminate whether V survives as Lefty, and indeterminate whether V survives as Righty. The positive case for this view may be summed up as follows.

Lefty and Righty, after fission, have separate bodies, brains and mental lives. (4JA) we should regard them as determinately distinct persons. Since $\text{Lefty} \neq \text{Righty}$, by the transitivity of identity, either $V \neq \text{Lefty}$ or $V \neq \text{Righty}$. So either: (4JB) both $V \neq \text{Lefty}$ and $V \neq \text{Righty}$; or (4JC) either $V = \text{Lefty}$ or $V = \text{Righty}$.

Suppose we endorse 4JB. It is implausible to regard V as continuing to exist as someone other than Lefty or Righty. Therefore, on 4JB, fission terminates V's existence. But this is an unattractive view, for reasons given in Subsection 4.3.5. It is more natural to think that (determinately) V survives fission. Therefore (4JC) (determinately) either $V = \text{Lefty}$ or $V = \text{Righty}$.

The Simple View is implausible. There is no further fact of DPI: whether $V = \text{Lefty}$ or $V = \text{Righty}$ depends wholly on the particular physical and psychological facts. But all these facts are symmetrical. So the relevant facts do not determine that $V = \text{Lefty}$ and they do not determine that $V = \text{Righty}$. But, determinately, V is either Lefty or Righty.

If the facts determined that $V \neq \text{Lefty}$, then, by disjunctive syllogism, from 4JC, they would determine that $V = \text{Righty}$.⁶⁹ But they do not determine that $V = \text{Righty}$. Therefore, it is not determinate that $V \neq \text{Lefty}$. Similar reasoning establishes that it is not determinate that $V \neq \text{Righty}$. Since, furthermore, it is not determinate that $V = \text{Lefty}$ and it is not determinate that $V = \text{Righty}$, it is indeterminate whether $V = \text{Lefty}$, and it is indeterminate whether $V = \text{Righty}$.

This conclusion makes sense of the natural thought that an object, when it undergoes (symmetrical) fission, in some sense "becomes two", having *undergone a process* of division. It "becomes two" in the sense that, after its division, there are two distinct items, such that it is indeterminate which of the objects it is identical with. It undergoes a process in that it exists both at the beginning and at the end of that event *via* which it "becomes two".

⁶⁹ Presumably, if both (a) it is determinate that p , and (b) from ' p ' *via* valid rules of inference we can derive ' q ', then (c) it is determinate that q . Cf. Subsection 2.5.1.

4.4 GARRETT'S CRITIQUE OF THE INDETERMINISTIC RESPONSE TO FISSION

4.4.1

I will now consider a critique of the indeterministic response to the Fission Case that is due to Garrett (1998: 64-67). This critique reappears in Garrett (2004). But (1998) will be my principle source.

Garrett argues that an indeterministic response has four flaws. I will address in turn each of the four alleged flaws.

4.4.2 Alleged lack of motivation

Garrett argues as follows that an indeterministic response to the Fission Case 'lacks ... motivation'. (1998: 64) I have added a label for each statement or question in the following argument, which I have labelled 'Argument GA'.

[Argument GA]

[4AA] It is determinate that Lefty, Righty and myself [*i.e.* V] are all persons. [4AB] It is perfectly determinate which relations of physical and psychological continuity we stand in to each other. [4AC*] Where is the logical space for indeterminacy? [4AD] Cases of alleged indeterminacy in identity over time typically arise when something is missing or diminished (as in the thought-experiment *Indeterminacy*). [4AE] In [the Fission Case], everything is present, twice-over. [4AC] There is no room for indeterminacy. (1998: 64)

Note that the thought-experiment Garrett calls '*Indeterminacy*' is the consideration of just such an operation as one of the central operations in the Combined Spectrum.⁷⁰

Certainly we should agree with 4AA. Each of Lefty, Righty and V is determinately a person. And in Fission, unlike in *Indeterminacy*, it is not obvious that there are candidate referents for any of these singular terms such that it is indeterminate whether that candidate is a person. So, indeterminacy regarding *personhood* does not seem, in the Fission Case, to provide 'room for indeterminacy'.

⁷⁰ Garrett's description of the thought experiment '*Indeterminacy*': 'An alteration machine changes me physically and psychologically. My brain is refigured so that roughly half of my memories, beliefs, (footnote continued next page)

4AB is also indisputable. The relations of physical and psychological continuity are all fixed by the description of the case.

Consider now 4AD. This seems the nub of Argument GA. In *Indeterminacy*, which is probably the case for which an indeterministic response is least controversial, the continuity relations hold to an intermediate – ‘diminished’ – degree. That this is so makes it reasonable to say of *Indeterminacy* that it is indeterminate whether the pre-operation person is the post-operation person. If the relations held to a high degree, the SDPI would be determinate (and true). If they held to a very low degree or not at all, the SDPI would again be determinate (and false). Given the object-candidate account, the diminished degree to which the continuity relations hold justifies the assertion that it is indeterminate whether the candidate referent which spans the operation is a person. But even without the object-candidate account, the diminished degree to which the continuity relation holds seems to justify an indeterministic response. In the Fission Case, however, the continuity relations between Lefty and V, and between Righty and V, are diminished in no significant respect. Psychological continuity is virtually perfect, since V’s brain is symmetrical and has the appropriate quantity and kind of redundancies. And physical continuity is not diminished in any *relevant* sense. In terms of what is relevant for questions of personal identity in relation to *e.g.* Lefty and V, the fact that only half of V’s brain makes it into Lefty’s skull is no more relevant than had a large piece of useless fat been shorn from V’s brain during the course of its journey to Lefty. Of course, the fact that the right half of V’s brain is brain and not useless fat is what enables the conundrum of Righty, and in that sense it *is* relevant; but it is not relevant to the question of relevant physical continuity between Lefty and V. And the failure of Lefty to share any of V’s pre-operative (and non-brain) body parts is just as irrelevant.

It would seem, then, futile to look for ‘room for indeterminacy’ by contesting 4AA, 4AB or 4AD.

Nevertheless, an indeterministic response to Fission is motivated in that the relation between the facts in fission – *i.e.* the problem’s ontological structure – and the logical constraints associated with the expressions in the SDPIs about which we want answers – *i.e.* the problem’s logical structure – shares a feature with the analogous

desires, and character traits are replaced with new and very different ones. It is vague or indeterminate whether I am psychologically continuous with the resulting person.’ (1998: 17)

relation in cases to which it is undoubtedly correct to respond by alleging indeterminacy. I call this feature ‘logical-ontological mismatch’. I will argue that logical-ontological mismatch, as an account of what justifies allegations of indeterminacy in particular cases, has more general application than an appeal to something being ‘missing or diminished’. Of course, Garrett does not suggest that diminishment or missingness justifies allegations of indeterminacy quite generally, but just that these features tend to justify such allegations in relation to cases pertaining to diachronic identity. However, I will argue that the Fission Case shares relevant structural features with cases pertaining to non-diachronic identity for which it is plausible to allege indeterminacy, despite the fact that such cases (like Fission) do not involve diminishment or missingness. Such cases do, however, like (at least) very many cases in which alleging indeterminacy is plausible, involve logical-ontological mismatch.

Consider first of all a paradigm case of indeterminacy. It is indeterminate whether Harry is tall. Logically speaking, ‘tall’ is a monadic predicate. Ontologically speaking, there is no natural boundary between tall persons and non-tall persons (or buildings, or trees, *etc.*). The natural structure of the facts relevant to tallness is rather continual and linear (with marginal qualifications relating to whether people have frizzy hair or are standing upright). Such continual and linear ontological structures are comparatively well-matched by predicates which are dyadic (and exhibit a transitive, asymmetrical structure). It is thus understandable why most actual ordered pairs of things (of some common kind) that are candidate ordered pairs for ‘is taller than’ are such as either determinately to satisfy that dyadic predicate or determinately to fail to satisfy it. There is a pretty sharp, and *non-obscure* natural boundary between the ordered pairs that are satisfactory in that respect, and the ordered pairs that are not so satisfactory. But there is no non-obscure natural boundary between the tall and the non-tall. Averaging all heights of all things of some kind might yield a natural boundary of sorts for ‘tall’ for that kind; but it would be an obscure one, requiring work to uncover. For ‘is taller than’ to yield a non-indeterminate predication when applied to the vast majority of ordered pairs of things of a common kind (*e.g.* two people, two buildings), that dyadic predicate needs to be burdened with but a small load of semantic information (in a sense to be elucidated shortly). For much of the meaning of the predicate is provided by the natural boundary, which can be “consulted” by speakers. There is a clear natural boundary between candidate ordered

pairs of people, buildings *etc.* of form $\langle x, y \rangle$ and their reversals of form $\langle y, x \rangle$. That ontological structure, so suited to carrying semantic information for 'is taller than', is not suited to carrying semantic information for 'tall', because no clear natural boundary is available for ready consultation in adjudicating between the satisfiers of 'is tall' and the non-satisfiers. Instead, speakers wishing to agree on a common boundary for 'tall' would have to consult a remembered or recorded piece of information, such as 'over 2 metres counts as 'tall' for a person'. It is in this sense that the word 'tall' would have then be burdened with a comparatively large load of semantic information. But it would be rather difficult for a community to agree on, and all remember, this common *artificial* boundary (which explains why no community has done so). So the semantic load required for a high degree of precision is too large to be worth carrying – explaining the vagueness of 'is tall'. In contrast, the *natural* boundary for 'is taller than' jumps out at everyone. The natural boundary to be found amongst the facts relevant both to 'tall' and to 'is taller than' is a boundary between one lot of *ordered pairs* and another; and this is due to the continual and linear ontological structure of these facts. No (non-obscure) natural boundary in amongst these facts divides two lots of *individuals* one from the other. So we can see how the fact that the ontological structure matches the dyadic structure but not the monadic one explains why the monadic one is very much vaguer, justifying allegations of indeterminacy regarding very many more predications of the monadic one than predications of the dyadic one.

Consider now another example, perhaps not quite so uncontroversially indeterminacy-involving as 'tall' and 'is taller than', but nevertheless very plausibly regardable as indeterminacy-involving. Suppose I am exploring a newly discovered continent. I come across a river-system: two rivers join; one river flows to the sea. I map the river system accurately, using thicker or thinner blue lines in accordance with the width of watercourses mapped. The resulting map looks like a wobbly but pretty symmetrical (by map standards) letter 'Y', but with a blob at the bottom to represent the sea, and something at the top to represent some mountains. I have named no geographical features at all, but have merely mapped. I return home and show my map to the King of my native country. I ask him, 'How many rivers are depicted in this map?'

I suggest that, if the King is like most people, he would *not* answer 'Three.' He would, I suggest, perhaps after pausing for thought, answer 'Two.'

Suppose now instead that I have mapped a river system in which one river splits, forming a virtually symmetrical delta of two branches, which branches flow completely separately, and for some time, before reaching the sea. I suggest that when shown an unlabelled map of *this* system, and asked how many rivers it depicts, the King would *not* answer ‘Three.’ He would, I suggest, answer either ‘One’ or ‘Two’, or, ‘That is a very strange question which I am unsure how to answer.’

Let us return now to the case of the tributaries, forgetting for now about the delta. It is very plausible to think the King right in answering ‘Two’. Were the line running from the top-left of the map to the bottom-middle very thick, and the line running from the top-right of the map to the centre (and joining the thick line) very thin, the King would certainly be wrong to say ‘Three’, and ought to answer ‘Two’. Flipping the map just described about its vertical axis again produces a map warranting only the answer ‘Two’. Between the two maps just described lies a spectrum of maps, in which the left branch of the ‘Y’ becomes progressively thinner, and the right branch progressively thicker, as we move from “left” to “right” along the spectrum. In all the maps, there is no significant difference in length between the two top bits of the ‘Y’. At all points in the spectrum the stem of the ‘Y’ (the bottom bit) depicts a single river. Pointing to a point on the stem, the King might say, ‘I dub this river ‘the River Murrie’.’ I, his cartographer-general, might label it thus:

R. Murrie → Y

Imagine the spectrum of maps each labelled as above, but such that each also depicts *another* river, to the right of, but totally unattached to, the first river system. Thus:

R. Murrie → Y I

I say to the King, in each case in this “Map Spectrum”, ‘Sire, the map is yet to be fully labelled. Please finish the job.’ In finishing labelling the map, there would be *no* case in the Map Spectrum in which *also* dubbing what the ‘I’ depicts ‘the River Murrie’ would be appropriate. I would be justified in saying, then, ‘Sire, you have already dubbed a *different* river ‘the River Murrie’. Calling two *different* rivers by the

same name will only lead to confusion.’ In contrast, suppose, in an arbitrary case in the Map Spectrum, that the King instructs me to label the top left bit of the ‘Y’ ‘*R. Murrie*’. I suggest, there would be *no* case in the Spectrum in which such labelling would be inappropriate, *except* for some of those cases in which the top-left bit of the ‘Y’ is clearly thinner than the top-right bit. In particular, in the central, symmetrical case it is still not inappropriate to label in this way. Since the top-left bit is not privileged in any relevant way over the top-right bit, it would be just as acceptable, in the central case, to label the top-right bit ‘*R. Murrie*’ (so dubbing what that bit represents).

An aside: In following a river upstream, explorers do not take themselves to reach the start of the river when they find it branching into equally powerful tributaries. This explains labels such as ‘The *Blue Nile*’ and ‘The *White Nile*’.

Consider now the Map Spectrum. Each map near the left determinately depicts just two rivers.⁷¹ The same goes for each map near the right. Now either:

(4MA) in each map in the spectrum, it is determinate that just two rivers are depicted;

or

(4MB) in some central map(s) in the spectrum, it is not the case that it is determinate that just two rivers are depicted.

Suppose 4MA holds. Consider a central case in the spectrum. Suppose just the bottom bit of the ‘Y’ has been labelled ‘*R. Murrie*’, thus dubbing what it represents. Presumably, the top-left river (*i.e.* whatever river is represented by the top-left bit of the ‘Y’) is not identical to the top-right river. Further, since the system is virtually symmetrical, it is highly implausible to suppose that it is determinate that the top-left river is numerically identical to the *Murrie*; and it is just as implausible to suppose this of the top-right river. Given that the top-left river is not identical to the top-right river, given 4MA, and given that the *Murrie* is a river, it follows that the *Murrie* is

⁷¹ The spectrum is really a spectrum of *cases*, in each of which I return from exploring a river-system that is *accurately* mapped by the relevant map.

either the top-left river or the top-right river, but it is indeterminate whether it is the top-left river and it is indeterminate whether it is the top-right river. So the statement 'The Murrie is numerically identical to the top-left river' is indeterminate. By addition of a disjunct:

- (4MM*) Either 'The Murrie = the top-left river' is indeterminate, or
'The Murrie = the top-right river' is indeterminate.

Suppose now that 4MB holds. Consider the middle map. If 4MB holds, at least this middle map verifies 4MB. So it is not the case that it is determinate that the middle map depicts just two rivers. Either determinately the middle map depicts more than two rivers, or it is indeterminate whether the middle map depicts two rivers. I suggest that

- (4MC) it is not the case that it is determinate that the middle map depicts more than two rivers.

I suggest that it is at very least *acceptable* to say that the middle map depicts two rivers. This explains why it would be acceptable to label both the top-left bit and the bottom-middle bit 'R. Murrie', thereby dubbing a river running all the way from the top-left to the bottom-middle 'R. Murrie'; and that the acceptability of this is non-trivial is demonstrated by the unacceptability of labelling both the bottom-middle bit of the 'Y', and the separate 'I', 'R. Murrie'. So we have good reason to deny that it is determinate that the middle map depicts more than two rivers. So, given the assumption 4MB, it is indeterminate whether the middle map depicts two rivers. Determinately, the top-left river is not the top-right river. So (in the system depicted by the middle map) there are determinately at least two rivers. If both (4MD) the Murrie is determinately not the top-left river and (4ME) the Murrie is determinately not the top-right river, then both (4MF) there are determinately three rivers, and (therefore) (4MG) it is determinate that there are not two rivers. But it is indeterminate whether there are two rivers. So

- (4MH) either: it is not the case that the Murrie is determinately not the top-left river; or it is not the case that the Murrie is determinately not the top-right river.

Suppose (4MI) the Murrie determinately *is* the top-left river. Then, since the top-left river is determinately not the top-right river, (4MJ) the Murrie is determinately not the top-right river. But the conjunction of 4MI and 4MJ is very implausible. So (4MK) it is not the case that the Murrie is determinately the top-left river. By similar reasoning, (4ML) it is not the case that the Murrie is determinately the top-right river. With 4MH, this allows us to conclude that

- (4MM) either 'The Murrie = the top-left river' is indeterminate, or 'The Murrie = the top-right river' is indeterminate.

4MM is just 4MM*. (Recall that '*S*' is indeterminate if and only if it is not the case that determinately *S* and it is not the case that determinately not-*S*; see Subsection 1.2.4.)

We have seen, then, that 4MA implies 4MM; and so does 4MB. And either 4MA holds, or 4MB does. All this reasoning has concerned the middle case of the Map Spectrum. We may conclude that the middle case of the Map Spectrum gives rise to indeterminate identity statements. Call this case 'Middle Map'.

What explains the indeterminacy surrounding Middle Map? There seems to be nothing 'missing or diminished' here.

An initial point to note is the structural similarity between Fission and Middle Map. Fission seems, indeed, more relevantly similar to Middle Map, which involves rivers joining, than to the case of the two-forked delta. Superficially, this might seem an odd assertion, given the "direction of flow". But consider that, in the case of the delta, it seems perhaps acceptable to assert that there is but *one* river depicted; whereas the analogous claim of the Fission Case is hard to swallow.

Having noted the structural similarity between Middle Map and Fission, and having noticed that neither involves diminishment or missingness, one wonders why, given that the former yields indeterminate identity statements, the latter should be taken not to do so. Fission is *structurally* more similar to Middle Map than to *Indeterminacy* (a central case in the Combined Spectrum). That Fission involves

persons and *diachronic* identity are the respects in which Fission is more like *Indeterminacy* than Middle Map. But given that we accept indeterminate SDPIs in, e.g., the Combined Spectrum, why should this consideration be significant? It seems that when it comes to *motivating* indeterminacy, “structural” matters are more important than whether persons or diachronic identity are involved. It is a mismatch between logical and ontological structure that seems to underpin indeterminacy. Let us see now how such a mismatch gives rise to the indeterminacy associated with Middle Map.

The logical structure of the numerical identity relation is familiar: a reflexive, transitive and symmetric relation. What ontological structure does this logical structure best match? It matches best domains of items without extension, e.g. mathematical objects. These objects do not threaten even to touch one another. But items with extension in space or time can be awkward for identity.

To see this, consider a domain of extended items which *often* touch one another or overlap, e.g. clouds or puddles. In a domain of clouds each relatum is a cloud. But since each relatum is extended, I might point to some particular part of it. Then if I dub the relatum I point to, I dub a cloud. But the extendedness of the cloud allows that I might dub either pointing here, or pointing there. A cloud needs a certain togetherness. Dissipated, or connected by a thin wisp of vapour, we wonder whether there is one cloud or two. This again arises from spatio-temporal extendedness. Thus when I dub cloud *a* pointing here, and then cloud *b* pointing there, when the places I point to are in vapours connected but slightly, it can be indeterminate whether $a=b$. What in the ontological structure (more specifically than the extendedness) gives rise to this? A cloud is a bunch of water-vapour molecules varying distances apart. The ontological structure is most precisely described by giving the position of each molecule. There is no natural boundary in a spectrum consisting of a bunch of molecules becoming progressively more distant from one another. There could only be an artificial boundary. But the informational burden of any sharply specified such artificial boundary is too great to be carried by any such predicates as ‘ $\lambda x \lambda y$ (vapour molecule *x* is in the same cloud as vapour molecule *y*)’.

Consider now the case of rivers. A river system has a branching ontological structure; and the extendedness of rivers allows us to dub a river by pointing at it here, or to dub a river by pointing at it there. The meaning of ‘river’ allows a single river to exist in two places, between which places other rivers may join it. Usually rivers of

differing strengths join. When a tiny river x runs into huge one y , the tiny river terminates. It is not the case that y is numerically identical to x . A rule develops: a river is generally said to terminate when it runs into a significantly broader or stronger body of water. This rule usually allows the individuation of rivers, thus allowing the branching ontological structure to cooperate with the non-branching structure of numerical identity. That this rule facilitates easy individuation *much* of the time is due to the fact that ‘significantly broader or stronger’ is clearly satisfied, or clearly not satisfied, by most actual ordered pairs of rivers. But I suggest that no widely accepted rule says that a river meeting an *equally* mighty river terminates, and is at the same time so precise as to rule out indeterminacy for any statement of synchronic river identity. Why? Such a rule would have virtually no actual use if it applied only in cases of exact symmetry in all respects. Not even a simple rule is worth the informational cost of retaining if it is virtually unusable. So if such a rule were really intrinsic to the meaning of ‘same river’, it would not require exact symmetry. Then a river would be said to terminate when it met a river of *around* the same breadth and strength as it. But then *such* a rule, owing to the many degrees of breadth and strength, would have no informationally cheap way of being carried in a form that ruled out indeterminacy in *all* cases. There would be asymmetrical ‘Y-systems’ where it would be indeterminate whether there are two or three rivers. So, logical-ontological mismatch again would lead to a justifiable claim of indeterminacy. So if there were a “tie-terminating” rule with sufficiently broad application to warrant the informational burden of carrying it, it would still not significantly decrease the number of cases involving indeterminacy. So there is thus little point carrying a tie-terminating rule. This explains why there does not appear to be any well-established tie-terminating rule⁷² for rivers. And this, in turn, explains why the dominant response to Middle Map does *not* involve asserting that the number of rivers depicted in it is determinately greater than two.

Some cases of indeterminacy-generating logical-ontological mismatch involve diminishment or missingness; but not all do. Perhaps those cases involving diachronic identity for which alleging indeterminacy is uncontroversial do involve diminishment or missingness. But why should this be a good reason to think the controversial cases

⁷² I do not mean ‘rule’ here to be synonymous with ‘application rule’ as defined in Chapter 2. A “well-established” rule would, I think, be a rule – though perhaps not an application rule – that almost
(footnote continued next page)

involving diachronic identity warrant allegations of indeterminacy *only* if diminishment or missingness is involved? Indeterminacy in general stems from something more general than diminishment or missingness. That general explanation also explains the most plausible response to Middle Map, which is structurally like the Fission Case. That general explanation involves logical-ontological mismatch.

It is hard to see why the temporal rather than spatial orientation of the structure in Fission should constitute a relevant difference with Middle Map.

Is diachronicity relevant because endurantism might be true, while the spatial analogue of endurantism is certainly not? No. Whether we are endurantists or perdurantists, we ought to agree that there can be indeterminate cases of diachronic identity, including for persons. Presuming the correctness of the object-candidate account of that indeterminacy – and I see no feasible alternative – the question of endurantism *versus* perdurantism is neither here nor there. At most, it is a challenge for the endurantist, not a problem for the object-candidate account.

In any case, there are predicates, such as ‘is the same puddle as’, for which an indeterministic approach to symmetrical *diachronic* cases of branching seems appropriate. Consider a large puddle, on uneven ground, gradually drying up, until there are two equally sized, shallower puddles. Is it really plausible to say that the original puddle has gone out of existence, merely because these two shallower puddles are around the same size?

Further, given that the case *Indeterminacy* yields indeterminate statements of diachronic personal identity, it is hard to see how the fact that *persons* are involved should make a difference.

To summarise: an indeterministic response to the Fission Case is motivated by the mismatch between the branching ontological structure, and the non-branching transitive and symmetrical logical structure of numerical identity.

4.4.3 Alleged counterintuitiveness

Garrett argues in the following way that an indeterministic response to Fission ‘is counterintuitive.’ (1998: 64)

everyone assumes almost everyone follows, in their usage of the expression to which the rule pertains.

[Argument GB]

[4BA] I exist prior to fission; [4BB] Lefty and Righty exist after fission. [4BC] If it is indeterminate whether I am Lefty, then it is indeterminate whether Lefty exists prior to fission. [4BD] The same is true of Righty. In which case, [4BE] it is indeterminate how many persons exist prior to fission. But, as our discussion of [the thesis that Lefty and Righty exist prior to fission, but only become spatially separate after fission] brought out, our common-sense intuition is that [4BF] there is, determinately, one and only one person who occupies the pre-fission body. (1998: 64)

4BE is certainly counterintuitive. But an indeterministic response to fission is not committed to it, and may deny it on just the grounds that it is counterintuitive. Below is a good argument against 4BE. In following this argument, recall that 'S' is determinately true if true on every admissible precisification; and that each admissible precisification will respect such logical norms as the transitivity of identity, and Excluded Middle. Also, we may presume that on no admissible precisification is Lefty = Righty.

Since Lefty is determinately not Righty, it is determinate that I (*i.e.* V) am not both Lefty and Righty. On some admissible precisification I am Lefty, on some other I am Righty, but on none am I both. If I am Lefty, Lefty exists prior to fission but Righty does not, and so just one person exists prior to fission (and just two persons exist after fission). If I am Righty, Righty exists prior to fission but Lefty does not, and so just one person exists prior to fission (and again just two persons exist after fission). If I am neither Lefty nor Righty, neither Lefty nor Righty exists prior to fission, but I do; and so again, just one person exists prior to Fission. Since (4N) both (4N*) it is determinate that I am either Lefty, or Righty, or neither Lefty nor Righty, and (4N**) it is determinate that I am not both Lefty and Righty, (4O) determinately just one person exists prior to fission. Therefore it is not the case that (4BE) it is indeterminate how many persons exist prior to fission.

The important thing to notice about the above argument is that none of its steps conflicts with the indeterministic response. In particular, as we saw in Subsection 2.5.1, a disjunction can be determinately true though none of its disjuncts is determinately true. So 4N* does not rule out the indeterministic response. 4N* is a classical tautology, and is still one under Anti-Determinist Classical Bivalentism (and, for that matter supervaluationism). There is no reason a defender of an indeterministic

response should be expected either (a) to deny that, determinately, numerical identity is transitive, or (b) to deny that Lefty is determinately non-identical with Righty.

In summary, the case for rejecting the counterintuitive 4BE, and affirming the plausible 4BF, can be wielded comfortably by one defending an indeterministic response to Fission.

Note also that the indeterministic response can comfortably affirm that the number of persons existent *after* fission is determinately two. It does *not* hold that V exists then in *addition* to Lefty and Righty. Rather, on one admissible precisification, it is settled that $V=Lefty$ and $V\neq Righty$ – so that there are just two persons – and on the only other admissible precisification, it is settled that $V=Righty$ and $V\neq Lefty$ – again, just two persons.

Thus, the plausible version of the indeterministic response, like the orthodox best candidate theory, respects our synchronic concept of a person, since it asserts that in the Fission Case there is, *simpliciter*, at each point in time, just one person per body.

4.4.4 An unorthodox version of the best candidate theory

The *best candidate theory*, as applied to fission cases, says (put roughly) that, when there are multiple candidates for identity with V, the *best* candidate is identical with V. The *locus classicus* of this approach is probably Nozick (1981). The relevant dimensions of “bestness” might include various factors, depending on one’s preferred view, but will typically include such things as degree of physical and/or psychological continuity. For example, had the left hemisphere but not the right one been erased of most memories prior to the transplant, then one might say that Righty is now the better candidate, so that $Righty=V$ and $Lefty\neq V$. According to *orthodox* best candidate theory, if there is no unique best candidate in a forward-branching case, the original person terminates at the point of fission. In the Fission Case, since Lefty and Righty are qualitatively alike in all relevant respects, orthodox best candidate theory says that V terminates at fission, as we have seen in Subsection 4.3.5.

One criticism sometimes made of the orthodox best candidate theory is that it makes personal identity – or at least, personal existence (see Garrett 1998: 67-70) – depend on “extrinsic” factors. Noonan (1989, Chs. 7 & 12), for example, makes much of this criticism. For Noonan:

if two events are parts of the history of a single entity of a kind in one situation then they must also be parts of the history of a single entity of the kind in any second situation in which, as judged by the Cambridge criterion, both they, and all the events which are parts of the history of the entity in the first situation, remain present. (1989: 139) (This passage is italicised by Noonan.)

We need not concern ourselves with the exact form of the ‘Cambridge criterion’ here; the nub of Noonan’s argument is this. Had V’s right hemisphere been destroyed, and only the left transplanted, a person would exist who is (determinately) identical with V. The history of the left hemisphere would then have traced the history of a single person. But the fact that the right hemisphere is also transplanted means that no such single person exists, on the standard best candidate theory. The history of the left hemisphere would in this case *not* trace the history of a single person. Noonan finds this odd, given that the same series of physical events – excision from the V-skull, insertion into the L-skull, and connection to the L-body – is involved in both cases, in relation to the left hemisphere. Noonan sees this as a highly unpalatable consequence of orthodox best candidate theory.

A variation of this critique concerns the ‘extrinsicness of existence-dependence’. (Garrett 1998: 70) Had the transplantation of the right hemisphere been unsuccessful, that hemisphere being destroyed in the operation, then the occupier of the L-body after the operation would have been V. But on orthodox best candidate theory, in the case of successful fission, a person occupies the L-body who is *not* V. As a consequence, after the successful fission, we can point to *Lefty* and say truly, ‘Had the transplant of the *right* hemisphere not succeeded, *this* person would not have existed.’ (Garrett 1998: 68) (This is not a direct quote.) This seems rather odd.

These consequences have been thought by some to make best candidate theory unacceptable. Garrett thinks that the extrinsicness of existence-dependence ‘may be surprising’ (70), but he argues (67-70) that it is ‘not objectionable’. I myself agree with Garrett on this point, although I will not repeat Garrett’s detailed defence of the best candidate theory here. I am more concerned here with the merits or demerits of *orthodox* best candidate theory as compared to the indeterministic response.

Garrett argues, as follows, that an indeterministic response to the Fission Case ‘is simply an unorthodox or non-standard version of the best candidate theory...’ (1998: 65)

[4XA] The [indeterministic] response holds that it is indeterminate whether I am Lefty, and indeterminate whether I am Righty. [4XB] That is, it is neither true that I am Lefty, nor true that I am Righty. [4XC] The reason for the indeterminacy must lie in some feature of the duplication (though as noted above, it is hard to see how duplication could induce indeterminacy).

However, [4XD] if Lefty had not existed, I would survive in the right-hand branch. There would then be no indeterminacy in my identity with the sole survivor. This implies that [4XE] whether or not it is true that I survive depends upon whether I have one off-shoot or two. (1998: 64-65)

Garrett concludes: Since the indeterministic response ‘is simply an unorthodox or non-standard version of the best candidate theory... why not simply opt for the standard version, and avoid the problems with the present response?’ (1998: 65)

Is this good reason to reject an indeterministic response?

Firstly, if it is indeterminate whether V is Lefty and indeterminate whether V is Righty, this does not entail that it is not true that V is Lefty, and it does not entail that it is not true that V is Righty. As we saw in Chapter 2, genuine indeterminacy is compatible with Bivalentism. Further, on a plausible indeterministic response, there are just *two* admissible precisifications of ‘same person’ in the fission case: (a) one on which Lefty = V; and (b) one on which Righty = V. Since V survives on all admissible precisifications, V determinately survives fission. So the indeterministic response is *not* committed to 4XE.

Having said this, it is still the case, on this indeterministic view, that whether (a) the history of the left hemisphere determinately traces the history of a single person, or (b) it is indeterminate whether the history of the left hemisphere traces the history of a single person, depends on factors “extrinsic” to the history of the left hemisphere.

But all this amounts to is this. The commitment of orthodox best candidate theory to taking extrinsic factors to be significant to such matters is *itself* no reason to reject orthodox best candidate theory in favour of an indeterministic response; since the latter is also committed to “extrinsicity”.

I do not propose here to meet the extrinsicity-based criticism of the best candidate theory. Garrett (1998: 67-70) has argued convincingly that this criticism has

little force. In defending orthodox best candidate theory, he argues that ‘the *extrinsicness of existence-dependence*’, though perhaps a ‘surprising’ result, ‘is not objectionable.’ (70) (Garrett’s emphasis.)

But let us ask: If the indeterministic response is a version of the best candidate theory, why should we prefer the orthodox version over the indeterministic one? Garrett’s argument here seems to be this: ‘why not simply opt for the standard version, and avoid the problems with the [indeterministic] response?’ (1998: 65)

We have already seen that the first two alleged problems, *viz.* lack of motivation and counterintuitiveness, are not problems at all. We shall shortly see that Garrett’s other criticism of the indeterministic response is not effective either. And we have seen in Section 4.3 that, unless there are telling objections to the indeterministic response, we ought to prefer it over orthodox best candidate theory.

Consider, too, that any plausible version of even the standard best candidate theory will admit indeterminacy in at least *some* cases of branching. Consider the following spectrum of cases, the “Branch Spectrum”. Each of these cases is like Fission, except in the following respects. In every one of these cases, Righty does not die till 50 years after the fission. In the “rightmost” case, Lefty dies when Righty dies. In the “leftmost” case, the brain hemisphere implanted into the L-skull remains viable only for one second after being implanted. Each case in the spectrum differs from the case to its left only in that the brain implanted into the L-skull remains viable for one second longer.

It is most implausible to deny that in the leftmost operation $V=Righty$. This case is not significantly different from the straightforward destruction of the left hemisphere and transplantation of the other. A plausible version of the best candidate theory should allow that at least one dimension of “bestness” (besides physical and/or psychological continuity) is whether the candidate in question is around for a significant length of time after the branching. In Nozick’s words: ‘It seems so unfair for a person to be doomed by an echo of his former self.’ (1981: 105-106.)

But then it is rather implausible to suppose that there is a sharp boundary between the cases in the Branch Spectrum in which $V=Righty$, and the cases in which $V\neq Righty$. Certainly in, *e.g.*, the second case from the right, where Lefty is around for just under 50 years after fission, it is implausible to suppose that $V=Righty$ here but $V\neq Righty$ in the rightmost case. So if the standard version of best candidate theory is correct – so that in the rightmost case, (it is determinate that) $V\neq Righty$ – there is no

sharp boundary between (a) the cases in which $V \neq \text{Lefty}$ and $V = \text{Righty}$, and (b) the cases in which $V \neq \text{Lefty}$ and $V \neq \text{Righty}$. By reasoning similar to that used in the Combined Spectrum Argument, there are some cases in which it is indeterminate whether $V = \text{Righty}$. Thus any plausible version of standard best candidate theory is committed to the postulation of indeterminacy in at least some cases of branching, even if it is not committed to such postulation in symmetrical fission.

Given that any plausible version of standard best candidate theory will allow for indeterminacy in some cases of branching, there is no grounds to prefer the standard version on the basis that it is simpler in virtue of affording a “neat” (indeterminacy-free) solution for all cases of branching.

It might be retorted that the indeterministic version of the best candidate theory is less simple because it affords a less “neat” response to the (symmetrical) Fission Case. However it is unclear how this constitutes simplicity. Looked at from another angle, the indeterministic best candidate theory (on the version according to which V determinately survives) seems simpler and more uniform than the standard best candidate theory, since V survives in all cases of branching. But the standard theory makes the question of whether V survives turn on whether the branching is lopsided.

It might perhaps be thought that the standard theory is to be preferred because it can be stated more simply. But I think that even if this were so, the brevity with which rival theories can be stated is quite a minor consideration compared to other reasons we have for preferring an indeterministic response. (See Section 4.3.)

Both versions of the best candidate theory can agree in endorsing the following plausible principle.

- (4G) If both (a) person y at time t^* is a candidate for identity with person x at earlier time t , and (b) for any z at t^* , y is a significantly better such candidate than z , then (c) $y=x$.

But only the standard version also endorses the following principle.

- (4H) If there exist at time t^* multiple candidates for identity with person x at earlier time t , so that for no y at t^* is y a better (*or*:

significantly better) candidate than any other candidate, then:
for no z at t^* is $z=x$; or in other words, x does not exist at t^* .

The indeterministic response can be construed as stipulating nothing about who is who in cases of multiple equally good candidates – even though it endorses 4G. The indeterministic account of who is who in branching cases⁷³ is somewhat like Horwich's (1997) incomplete definition of 'glub' discussed in Subsection 2.6.3. It does not specify in advance a correct answer for all cases. Rather it leaves some cases – cases of approximately symmetrical branching – open.

4.4.5 Insufficient generality?

Garrett's (1998: 64-67) fourth and final criticism of the indeterministic response to the Fission Case appears to be that it is insufficiently general. He argues that '[i]t fails to apply to scenarios which don't involve persons, but which are relevantly similar to *Fission*. In particular, [it] fails to apply to the story of the *Ship of Theseus*.' (1998: 64)

The modern puzzle of the ship of Theseus can be presented as a pair of cases.⁷⁴

In Case A, we begin with Theseus' ship. The planks composing it are removed and replaced one at a time, gradually over some period of time, with fresh planks, until none of the original planks remain. The resulting ship may be called 'the continuously repaired ship'. Suppose also that the removed planks 'are retained and used to build another ship, exactly similar to Theseus' ship. Call this 'the re-constituted ship'.' (Garrett 1998: 65)

Case B is just like Case A except in that the removed planks are not replaced. Call the only ship existing at the conclusion of this case 'the re-constituted ship_B'.

As Garrett argues, the intuitively correct response to these cases is that, in Case A, Theseus' ship = the continuously repaired ship (and Theseus' ship \neq the re-constituted ship), but in Case B, Theseus' ship = the re-constituted ship_B.

⁷³ I ignore back-branching (fusion) cases here. They would not seem to pose any special problem for an indeterministic account, though.

⁷⁴ For the story, Garrett credits: (a) Plutarch *Lives* (which Garrett quotes from sections 22-23, as cited at p. 92 n. 15 of Wiggins 1980 *Sameness and Substance* Oxford: Basil Blackwell); and (b) Hobbes, T. *De Corpore*, part II, ch. II, in Molesworth, ed. 1839-1845 *The English Works of Thomas Hobbes* London: John Bohn, vol. 1 (which Garrett quotes from p. 136, as cited at p. 92 of Wiggins 1980 *Sameness and Substance*).

Garrett argues that ‘the indeterminacy response cannot apply [in these cases] since, in [Case A], the ship of Theseus is definitely the continuously repaired ship and definitely not the re-constituted ship. There is no room for the indeterminacy response... Yet given the structural similarities between *Fission* and *Ship of Theseus*, it would be good if one response could cover them both. This is precisely what is offered by [orthodox?][best candidate theory].’ (1998: 67)

Certainly a generalisation of orthodox best candidate theory can cover both cases: in Case A, the continuously repaired ship is a significantly better candidate for identity with Theseus’ ship than the reconstituted ship. This would be just the answer yielded by the ship-related analogue of 4G, “4G_{ships}”:

- (4G_{ships}) If both (a) ship y at time t^* is a candidate for identity with ship x at earlier time t , and (b) for any z at t^* , y is a significantly better such candidate than z , then (c) $y=x$.

Similarly, 4G_{ships} would yield the answer that in Case B, Theseus ship = the re-constituted ship_B.

But a plausible (generalisation of the) indeterministic version of best candidate theory would also yield just these answers, for it can happily endorse 4G_{ships}, just as it can endorse 4G, as we have seen. It just does not endorse the further principle, 4H. And in order to yield the correct answers to Cases A and B, it need not endorse the ship-related analogue of 4H, “4H_{ships}”. Moreover, it need not endorse the implausible 4H_{puddles}; and nor need it endorse the implausible synchronic analogue of 4H for rivers (according to which, paddling upstream, one comes to the end of a river when one comes to a point symmetrical branching). That is not to say that standard best candidate theory is *committed* to 4H_{puddles} *etc.*. But if one is to compare the two versions of best candidate theory by comparing the plausibility of generalisations of them, it seems that, if anything, the indeterministic version comes out on top. For, as we have seen, we have good reason to deny, *e.g.*, that there are determinately *three* rivers in Middle Map.

Having said this, I do not think that a generalised version of *either* theory is necessarily plausible. Why should only a blanket approach to individuating objects capable of fissioning be a plausible one? I think there are many cases like Middle Map, that are awkward for generalised standard best candidate theory. But there may

be other cases in which (an analogue of) indeterministic best candidate theory is less plausible. What we are interested in here is the more reasonable response for cases of *personal* fission.

4.4.6

In summary, an indeterministic response to the Fission Case is coherent and plausible. Given the presumptive case for an indeterministic response outlined in Section 4.3, and given that the criticisms just examined are unconvincing, we should conclude that this is the best response to the case.

In particular, this response allows us both to respect the logic of numerical identity and to avoid distorting our concept of a person. Like the orthodox best candidate theory, it respects our *synchronic* concept of a person, since it claims that, determinately, exactly one person exists prior to fission, and exactly two persons exist after fission. But unlike the orthodox best candidate theory, it also accords with an intuitive *diachronic* conception of a person, according to which a person will survive if enough of his/her brain continues to exist and to support a continued mental life.

4.5 PRACTICAL CONSIDERATIONS

An interesting question is this: ‘What is it that we disagree about, in debating the question of *who is who* in the Fission Case?’

If two *Simple Theorists* have all the information provided in the description of the Fission Case, then, if they yet disagree about who is who, there is, for them, a fact of the matter about who is who, and they are disagreeing about what this fact is.

But consider a debate between two *Complex Theorists* regarding who is who in the Fission Case. Assuming they know all the physical and psychological facts, they are not disagreeing about any *other* fact: for they are Complex Theorists, and agree that there is no further fact.

It seems that their disagreement concerns the following question. If one has to use the language of personal identity to describe the case, what *ought* one to say about who is who, *given* the facts? Answering this question involves an investigation of the

criteria implicitly attached to predicates such as ‘person’. These criteria may not be obvious, but philosophical discussion may help to reveal them.

Having said this, even after the particular physical and psychological facts are all known, the question of what we *say* about what is a person, what is not, and who is who, is not *merely* verbal. For the words ‘person’ and ‘same person’ seem to have practical import. Normative concepts such as *desert* seem to be structured around these concepts.⁷⁵ Thus, to use the term ‘person’ is not merely to report physical and psychological facts. It also carries normative significance. To call something a person, or to call *x* and *y* the same person, is not *just* to utter an indicative sentence. It is also an *act*. For example, if a society holds a person responsible only for that *same* person’s past actions – if that connection between diachronic personal identity, and responsibility, is held constant – then to promote, say, the view that V is identical with Lefty is to promote the view that Lefty is responsible for V’s actions.

I would prefer to remain neutral over the question of whether the disagreement over what to say in the Fission Case is a disagreement about how we ought to *refine* or to *extend* the meaning of ‘person’, as opposed to a debate about which answer we are already implicitly committed to in virtue of our concepts as they *already* are. And even if the debate does, as Johnston argues, concern ‘how to extend our practice to a case where it presently gives no answer’ (1992: 603), pointing *this* out is not by itself a good argument for the claim that an indeterministic response to the Fission Case is the *best* response, *i.e.* the *best* way of extending (or leaving “unextended”) our linguistic practice here. Rather, further arguments – such as those given earlier in this chapter – are required to establish that this is the best option.

Those arguments so far considered have been more of a “formal” nature than of a practical one. We have seen, for example, that an indeterministic response allows a Complex Theorist to respect the logic of identity without distorting our synchronic or diachronic concept of a person. *Formal* considerations do seem important in deciding how to refine or to settle our concepts in relation to particular puzzle cases. However, there is no reason not to take *practical* considerations into account also, especially given the normative importance we seemingly attach to questions of who is identical to whom.

⁷⁵ That they should be so structured is at least the presumption. But see Parfit (1984: 323-326). I do not claim to address fully Parfit’s normative arguments here.

I do not propose here to do complete justice to the normative debate concerning “what matters in survival”. But it seems, at least *prima facie*, that practical considerations do more to support the view that V determinately survives fission, than to undermine that view. If we think that a person x no longer exists after a certain point, it seems natural to make certain normative claims. It seems natural to think that no-one other than x deserves punishment or reward for actions autonomously performed by x . But it would seem unfair to withhold rewards for V’s deserving actions from Lefty and Righty on account of the intervening fission. And it would seem wrong to absolve Lefty and Righty of punishment which is still due to V for V’s past actions – unless we did so out of sympathy for the arguably unfortunate situation that Lefty and Righty find themselves in.

Consider that the fission operation could have been performed without V’s, Lefty’s or Righty’s realising it, *e.g.* by anaesthetising V during his sleep, and taking Righty away to a secret location before Lefty wakes up. If the bodies of the three brothers are sufficiently similar, and the operation performed with sufficient expertise, Lefty and Righty would each believe he is V upon waking. (Righty would believe both that he is V and that he has been kidnapped during his sleep.) A week later, Righty could arrive on V’s doorstep, believing himself to have arrived at *his* home. Surely it would be wrong to deprive Lefty and Righty of the house of V, by exercising V’s will: they ought rather to continue, jointly, to own the house, or so it would seem. But V’s will would presumably be exercisable if V has ceased to exist.

Similarly, Righty and Lefty would both seem right, in this situation, in believing themselves entitled to receive whatever rewards were earned by V prior to fission, but not yet collected.

These practical considerations are not by themselves clinching arguments for the indeterministic interpretation. One might argue, *e.g.*, (4YA) that we ought to reject the normative intuitions described above regarding what Lefty and Righty deserve. Alternatively, one might argue (4YB) that our normative concepts are, or ought to be, structured around relations other than personal identity (see Parfit 1984 Chs. 12-15), relations like psychological and/or physical continuity, which hold between V and Lefty, and between V and Righty. One could then maintain that neither Lefty nor Righty is V, though each deserves what V would have deserved had he survived normally.

However, we should prefer, if we can, to retain intuitions about what Lefty and Righty deserve, rather than to abandon them. The indeterministic response allows us to retain these intuitions. We may hold, for example, that Lefty and Righty are *jointly* entitled to V's house, because V determinately still exists, and V is entitled to the house. That they ought to *share* the house seems natural given that neither Lefty nor Righty is *determinately* identical with V. Neither Lefty nor Righty deserves exclusive ownership of V's house.

So, we ought to reject 4YA. What about 4YB? This view would seem to entail that, since each of Lefty and Righty is psychologically and physically continuous with V, *each* deserves whatever V would have deserved had he survived normally. But suppose V has earned a \$10,000 reward from the local council, for having performed some brave action, but has yet to collect the reward when fission occurs. It would seem more natural to say that Lefty and Righty are jointly entitled to the \$10,000; than to say that the council ought now to disburse \$20,000 by giving one lot of \$10,000 to Lefty and another lot of \$10,000 to Righty. Only one \$10,000-warranting action has been performed; and so only one lot of \$10,000 has been earned. The indeterministic response yields just this outcome, since it is *V* who deserves the \$10,000, and V exists after the operation; that it is indeterminate *which* of Lefty and Righty is identical with V justifies their sharing the \$10,000 between them.

However, even if we accept 4YB, this does not justify rejecting the indeterministic response, since 4YB is still consistent with the view that V survives fission. If 4YB were true, practical considerations would weigh neither for nor against the indeterministic view. But we would still have other good reasons to prefer the indeterministic response, as detailed in Section 4.3.

We ought not to overstate the significance of these practical considerations for the question of who is who in the Fission Case. (Perhaps normative intuitions ought sometimes to be revised in the light of metaphysical conclusions.) But the practical considerations here seem to complement, and certainly do not conflict with, the conclusion to which we were earlier led by more formal considerations, *viz.* that V determinately survives fission, but that it is indeterminate whether he survives as Lefty, and indeterminate whether he survives as Righty.⁷⁶

⁷⁶ For an interesting discussion of the interplay between practical concerns, and questions of who is diachronically identical with whom, see Braddon-Mitchell and West (2001).

4.6 CONCLUSION

In the Fission Case: (a) V determinately survives the operation and exists after the operation; (b) determinately, V is either Lefty or Righty; (c) it is indeterminate whether V is Lefty; and (d) it is indeterminate whether V is Righty.⁷⁷

⁷⁷ For a view of fission that is somewhat in the spirit of mine, see Rieber (1998). While Johnston (1992) holds that it is indeterminate who is who in the Fission Case, he does not seem to claim (4CA) that V determinately survives and exists after fission; whereas I hold that the ability to endorse 4CA while denying the Simple View constitutes a powerful argument for an indeterministic response to the case.

CONCLUSION

Rather than listing again the main conclusions of each chapter, I will attempt here to summarise in quite general terms the picture of vagueness-related *indeterminacy*, and of *diachronic personal identity*, that I have tried to promote.

Firstly, we must distinguish between (a) variation between different possible worlds or situations and (b) variation between different possible representations of a single world or situation.

Failure to make this distinction with respect to *linguistic* representations of the world leaves one unable to make sense of what seems intuitively very plausible: that certain statements neither determinately hold, nor determinately fail to hold, on account of the vagueness of linguistic expressions in them. (See, in particular, Subsection 2.7.2.)

Failure to make this distinction in relation to the various *imaginative* pictures we might form of the world tempts one into believing the Simple View of diachronic personal identity, a view against which there is a strong presumptive case. (See, in particular, Subsection 3.4.3.)

Secondly, persons are not metaphysically privileged in such a way as to exempt statements of personal identity from being indeterminate. (See Chapters 1 and 3.)

Thirdly, a proper understanding of the nature of indeterminacy, coupled with an abandonment of any special reluctance to apply this concept to statements concerning persons, assists the investigation of the various puzzle cases upon which the philosophy of personal identity has often focused. In some such cases, to rule out an indeterministic response, through a misunderstanding of the nature of indeterminacy, or of its relation to personal identity, is to rule out a response which enables us to reconcile intuitions which, otherwise, would be irreconcilable. (See Chapter 4.)

BIBLIOGRAPHY

- Alter, T. & Rachels, S. 2004. 'Epistemicism and the combined spectrum'. *Ratio* 17: 241-255.
- Beck, S. 1998. 'Back to the self and the future'. *South African Journal of Philosophy* 17: 211-25.
- Braddon-Mitchell, D. & West, C. 2001. 'Temporal Phase Pluralism'. *Philosophy and Phenomenological Research* 62: 59-83.
- Burgess, J. A. 1998. 'In defence of an indeterminist theory of vagueness'. *Monist* 81: 233-252.
- Burns, L. C. 1991. *Vagueness: An Investigation into Natural Languages and the Sorites Paradox*. Dordrecht: Kluwer.
- Campbell, R. 1974. 'The sorites paradox'. *Philosophical Studies* 26: 175-191.
- Campbell, S. 2001. 'Is connectedness necessary to what matters in survival?' *Ratio (new series)* 14: 193-202.
- Cargile, J. 1969. 'The sorites paradox'. *British Journal for the Philosophy of Science* 20: 193-202. Reprinted in Keefe & Smith, eds. 1997: 89-98.
- Chibeni, S. S. 2004. 'Ontic vagueness in microphysics'. *Sorites* 15: 29-41. (Accessed Oct. 2005 at http://www.sorites.org/Issue_15/chibeni.htm.)
- Dorr, Cian. 2003. 'Vagueness without ignorance'. *Nous supplement: Philosophical Perspectives* 17: 83-113.

- Dummett, M. 1975. 'Wang's paradox'. *Synthese* 30: 301-324. Reprinted in Keefe & Smith, eds. 1997: 99-118. (I refer to page numbers from Keefe & Smith eds..)
- Edgington, D. 1997. 'Vagueness by degrees' in Keefe & Smith, eds. 1997: 294-316.
- Evans, G. 1978. 'Can there be vague objects?' *Analysis* 38: 208. Reprinted in Keefe & Smith, eds. 1997: 317.
- Field, H. 1994. 'Disquotational truth and factually defective discourse'. *The Philosophical Review* 103: 405-452.
- Field, H. 2000. 'Indeterminacy, degree of belief, and excluded middle'. *Nous* 34: 1-30.
- Field, H. 2003a. 'The semantic paradoxes and the paradoxes of vagueness', accessed Aug. 2005 at <http://philosophy.fas.nyu.edu/docs/IO/1158/Semantic.pdf>, pp. 1-39 (from which page numbers taken). (Also appears in Beall, J. C., ed. 2003. *Liars and Heaps*. New York: Oxford University Press. pp. 262-311.)
- Field, H. 2003b. 'No fact of the matter'. *Australasian Journal of Philosophy* 81(4): 457-480.
- Fine, K. 1975. 'Vagueness, truth and logic'. *Synthese* 30: 265-300. Reprinted in Keefe & Smith, eds. 1997: 119-150. (I refer to page numbers from Keefe & Smith eds..)
- Flew, A. 1984. *A Dictionary of Philosophy*. London: Pan Books Ltd.
- Garrett, B. 1998. *Personal Identity and Self-Consciousness*. London: Routledge.
- Garrett, B. 2004. 'Johnston on fission'. *Sorites* 15: 87-93. (Accessed Oct. 2005 at http://www.sorites.org/Issue_15/garrett.htm.)
- Gibbard, A. 1975. 'Contingent identity'. *Journal of Philosophical Logic* 4: 187-222.

- Goguen, J. A. 1969. 'The logic of inexact concepts'. *Synthese* 19: 325-373.
- Gómez-Torrente, M. 1997. 'Two problems for an epistemicist view of vagueness'. *Philosophical Issues* 8: 237-245.
- Goodenough, J. M. 1996. 'Parfit and the sorites paradox'. *Philosophical Studies* 83: 113-120.
- Hart, W. 1992. 'Hat-tricks and heaps'. *Philosophical Studies* 33: 1-24.
- Heller, M. 1996. 'Against metaphysical vagueness', in *Philosophical Perspectives* 10: 177-183.
- Horwich, P. 2000. 'Stephen Schiffer's theory of vagueness'. *Philosophical Issues* 10: 271-281.
- Hyde, D. 1994. 'Why higher-order vagueness is a pseudo-problem'. *Mind (Symposium: Higher-Order Vagueness)* 103: 35-41.
- Hyde, D. 1995. 'Commentary: proliferating conceptions of truth: comments on McGee and McLaughlin'. *Southern Journal of Philosophy* 33(Supp.): 253-261.
- Hyde, D. 1997. 'From heaps and gaps to heaps of gluts'. *Mind* 106: 641-660.
- Johnston, M. 1987. 'Human beings'. *Journal of Philosophy* 84: 59-83.
- Johnston, M. 1992. 'Reasons and reductionism'. *Philosophical Review* 101: 589-618.
- Johnston, M. 1997. 'Human concerns without superlative selves', in Dancy, J., ed. *Reading Parfit*. Oxford: Blackwell. pp. 149-179. Reprinted in Martin & Barresi, eds. 2003: 260-291. (I refer to page numbers in Martin & Barresi eds.)
- Keefe, R. 1995. 'Contingent identity and vague identity'. *Analysis* 55: 183-187.

- Keefe, R. & Smith, P., eds. 1997. *Vagueness: A Reader*. Cambridge, Massachusetts: MIT Press.
- Keefe, R. & Smith, P. 1997. 'Introduction: theories of vagueness', in Keefe & Smith, eds. 1997: 1-57.
- Korsgaard, C. M. 1989. 'Personal identity and the unity of agency: A Kantian response to Parfit'. *Philosophy and Public Affairs*. 18: 101-132.
- Lewis, D. K. 1970. 'General semantics'. *Synthese* 22: 18-67.
- Lewis, D. K. 1983. 'Survival and identity' and 'Postscript', in Lewis, D. *Philosophical Papers*, vol. 1, Oxford: Oxford University Press. pp. 55-70 & 73-77. Reprinted in Martin & Barresi, eds., 2003: 144-167. (Page numbers taken from Martin & Barresi eds.)
- Lewis, D. K. 1988. 'Vague identity: Evans misunderstood'. *Analysis* 48: 128-130.
- Lowe, E. J. 1994. 'Vague identity and quantum indeterminacy'. *Analysis* 54: 110-114.
- McGee, V. 1985. 'A counterexample to modus ponens'. *Journal of Philosophy* 82: 462-471.
- McGee, V. & McLaughlin, B. 1995. 'Distinctions without a difference'. *Southern Journal of Philosophy* 33(supp.): 203-252.
- McGee, V. & McLaughlin, B. 2004. 'Logical commitment and semantic determinacy: a reply to Williamson'. *Linguistics and Philosophy* 27: 123-136.
- Machina, K. F. 1976. 'Truth, belief and vagueness'. *Journal of Philosophical Logic* 5: 47-78.
- McLaughlin, B. 1997. 'Supervenience, vagueness and determination'. *Philosophical Perspectives* 11: 209-230.

- Madell, G. 1981. *The Identity of the Self*. Edinburgh: Edinburgh University Press.
- Madell, G. 1985. 'Derek Parfit and Greta Garbo'. *Analysis* 45: 105-109.
- Martin, R. & Barresi, J., eds. 2003. *Personal Identity*. Malden, MA: Blackwell Publishing Ltd.
- Martin, R. 1995. 'Fission rejuvenation'. *Philosophical Studies* 80: 17-40. Reprinted in Martin & Barresi, eds. 2003: 216-237.
- Mehlberg, H. 1958. *The Reach of Science*. Toronto: University of Toronto Press.
- Merricks, T. 2001. 'Realism about personal identity over time'. *Nous supplement: Philosophical Perspectives* 15: 173-187.
- Meyer, R. K., Routley, R. & Dunn, J. M. 1979. 'Curry's paradox'. *Analysis* 39: 124-128.
- Nagel, T. 1986. *The View from Nowhere*. New York: Oxford University Press.
- Noonan, H. W. 1989. *Personal Identity*. 2nd ed. London: Routledge.
- Nozick, R. 1981. 'Personal identity through time', in Nozick, R. *Philosophical Explanations* Cambridge, Mass.: the Belknap Press, imprint of Harvard University Press, 1981. pp. 29-48, 50-51, 58-61, 69, reprinted in Martin & Barresi, eds. 2003: 92-114. (Page numbers cited are from Martin & Barresi, eds.)
- Olson, E. T. 2002. 'Personal Identity', in Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy (Fall 2002 Edition)*. Accessed Aug. 2005 at <http://plato.stanford.edu/archives/fall2002/entries/identity-personal/>.
- Parfit, D. 1971a. 'On the importance of self-identity'. *Journal of Philosophy* 68: 683-690.

Parfit, D. 1971b. 'Personal identity'. *The Philosophical Review* 80: 3-27.

Parfit, D. 1984. *Reasons and Persons*. Oxford: Clarendon Press.

Parfit, D. 1986. 'Comments'. *Ethics* 96: 832-872.

Parsons, T. & Woodruff, P. 1995. 'Worldly indeterminacy of identity'. *Proceedings of the Aristotelian Society* 95: 171-91. Reprinted in Keefe & Smith, eds. 1997: 321-37. (I refer to page numbers from Keefe & Smith eds..)

Perry, J., ed. 1975. *Personal Identity*. Berkeley, CA: University of California Press.

Rieber, S. 1998. 'The concept of personal identity'. *Philosophy and Phenomenological Research* 58: 581-594.

Rorty, A. O., ed. 1976. *The Identities of Persons*. Berkeley, CA: University of California Press.

Sainsbury, R. M. 1990. 'Concepts without boundaries'. Inaugural lecture publ. by the King's College London Department of Philosophy. Reprinted in Keefe & Smith, eds. 1997: 251-264.

Salmon, N. U. 1981. *Reference and Essence*. Oxford: Blackwell.

Schiffer, S. 1998. 'Two issues of vagueness'. *The Monist* 81: 193-214.

Schiffer, S. 1999. 'The epistemic theory of vagueness'. *Philosophical Perspectives* 13: 481-503.

Schiffer, S. 2000. 'Vagueness and partial belief'. *Philosophical Issues* 10: 220-257.

Seibt, J. 2000. 'Fission, sameness and survival: Parfit's Branch Line argument revisited'. *Metaphysica* 1: 95-134.

- Shoemaker, S. 1963. *Self-Knowledge and Self-Identity*. Ithaca, NY: Cornell University Press.
- Shoemaker, S. 1970. 'Persons and their pasts'. *American Philosophical Quarterly* 7: 269-285.
- Shoemaker, S. 1984. 'Personal identity: a materialist's account' and 'Sydney Shoemaker's reply', in Shoemaker, S. & Swinburne, R. *Personal Identity*. Oxford: Basil Blackwell. pp. 67-132 & 139-152.
- Shoemaker, S. 2002. 'The irrelevance/incoherence of non-reductivism about personal identity'. *Philo* 5: 143-160.
- Sider, T. 1996. 'All the world's a stage'. *Australasian Journal of Philosophy* 74: 433-453.
- Sider, T. 2001. 'Criteria of personal identity and the limits of conceptual analysis'. *Philosophical Perspectives* 15: 189-209.
- Simons, P. 1992. 'Vagueness and ignorance'. *Proceedings of the Aristotelian Society*. 66: 163-177.
- Sorensen, R. 1988. *Blindspots*. Oxford: Clarendon Press.
- Sorensen, R. 2001a. *Vagueness and Contradiction*. Oxford: Clarendon Press.
- Sorensen, R. 2001b. 'Vagueness has no function in law'. *Legal Theory* 7: 385-415.
- Swinburne, R. 1974. 'Personal Identity'. *Proceedings of the Aristotelian Society* 74: 231-247.

Swinburne, R. 1984. 'Personal identity: the dualist theory' and 'Richard Swinburne's reply', in Shoemaker, S. & Swinburne, R. *Personal Identity*. Oxford: Basil Blackwell, pp. 1-66 & 133-138.

Tye, M. 1994. 'Sorites paradoxes and the semantic of vagueness'. *Philosophical Perspectives* 8: 189-206. Reprinted, minus §1, in Keefe & Smith, eds. 1997: 321-37. (I refer to page numbers from Keefe & Smith eds.)

Unger, P. K. 1979a. 'There are no ordinary things'. *Synthese* 41: 117-154.

Unger, P. K. 1979b. 'I do not exist', in Macdonald, G. F., ed.. *Perception and Identity*. London: Macmillan. pp. 235-251.

Unger, P. K. 1979c. 'Why there are no people'. *Midwest Studies in Philosophy* 4: 177-222.

Van Fraassen, B. C. 1969. 'Presuppositions, supervaluations and free logic', in Lambert, K., ed., *The Logical Way of Doing Things*. New Haven: Yale University Press. pp. 67-91.

White, S. 1989. 'Metapsychological relativism and the self'. *Journal of Philosophy* 86: 298-323.

Wiggins, D. 1967. *Identity and Spatio-Temporal Continuity*. Oxford: Basil Blackwell.

Wiggins, D. 1979. 'The concern to survive'. *Midwest Studies in Philosophy* 4: 417-422.

Williams, B. A. O. 1956-7. 'Personal identity and individuation'. *Proceedings of the Aristotelian Society* 57: 229-52.

Williams, B. A. O. 1970. 'The self and the future'. *Philosophical Review* 79: 161-180. (Reprinted in Williams 1973, and in Martin & Barresi, eds. 2003; but I cite page numbers from *Philosophical Review*.)

- Williams, B. A. O. 1973. *Problems of the Self*. Cambridge: Cambridge University Press.
- Williamson, T. 1992. 'Vagueness and ignorance'. *Proceedings of the Aristotelian Society* 66: 145-162.
- Williamson, T. 1994. *Vagueness*. London: Routledge.
- Williamson, T. 1996. 'Wright on the epistemic conception of vagueness'. *Analysis* 56: 39-45.
- Williamson, T. 1999. 'Schiffer on the epistemic theory of vagueness'. *Philosophical Perspectives* 13: 505-517.
- Williamson, T. 2004. 'Reply to McGee and McLaughlin'. *Linguistics and Philosophy* 27: 113-122.
- Wittgenstein, L. 1953. *Philosophical Investigations*, translated by G. E. M. Anscombe, 3rd edition, 1967, Oxford: Blackwell.
- Wolf, S. 1986. 'Self-interest and interest in Selves'. *Ethics* 96: 704-720.
- Wright, C. 1995. 'The epistemic conception of vagueness'. *Southern Journal of Philosophy* 33(Supp.): 133-159.
- Zadeh, L. A. 1965. 'Fuzzy sets'. *Information and Control* 8: 338-353. Reprinted at pp. 29-44 of Yager, R. R. et. al., eds. 1987. *Fuzzy Sets and Applications: Selected Papers by L. A. Zadeh*. New York: John Wiley.