

EL CONEXIONISMO Y SU IMPACTO EN LA FILOSOFÍA DE LA MENTE

Josep E. Corbí y Josep L. Prades

I. INTRODUCCIÓN

En sus inicios¹, a principios de los cuarenta, la investigación en Inteligencia Artificial responde a dos modelos fundamentales, a saber: el simbólico y el neural o conexionista. Por un lado, los modelos simbólicos intentan simular las capacidades cognitivas de los seres humanos mediante el procesamiento de fórmulas sintácticas, pues se entiende que tales capacidades descansan necesariamente en un sistema de representación cuyos elementos básicos reflejan la estructura sintáctica del lenguaje. Por otro lado, los modelos neurales o conexionistas encuentran en la estructura sináptica del cerebro la clave para simular la inteligencia humana y, por ello, tratan de imitar la actividad inteligente a partir de la elaboración de redes de conexiones entre unidades muy simples. La aparición en 1969 del libro de Marvin Minsky y Seymour Papert, *Perceptrons*, alteró este equilibrio al cuestionar la viabilidad de la estrategia conexionista.

Minsky y Papert argumentaron que los modelos neurales difícilmente podrían llegar a simular algún día habilidades cognitivas mínimamente complejas. Esta conclusión provocó al abandono de la rama conexionista de la Inteligencia Artificial, con lo que, desde entonces, todos los esfuerzos se centraron en los modelos simbólicos. Esta opción venía, además, avalada por datos procedentes de otras disciplinas como la lingüística, donde Noam Chomsky² reivindicaba la existencia de una sin-

1. Entre los trabajos más significativos de este periodo inicial se cuentan McCulloch y Pitts (1943), Hebb (1949), Neuman (1956), Rosenblatt (1959, 1962) y Selfridge (1959).

2. Cf. Chomsky (1957, 1959, 1968)

taxis universal e innata para dar cuenta de la sorprendente productividad de nuestra actuación lingüística. El impacto filosófico de esta maniobra excluyente en favor del paradigma simbólico fue más que notable, y dio lugar a lo que a lo largo de este capítulo denominaremos la *Imagen Sintáctica de la Mente*³.

Con todo, el acuerdo en torno a la exclusividad de los modelos simbólicos se quebró a principios de los años 80; pues, por una parte, los modelos simbólicos habían tropezado con dificultades importantes que empezaban a ahogar el entusiasmo inicial⁴ y, por otra, surgían nuevos y atractivos modelos conexionistas que cosechaban éxitos donde los modelos simbólicos parecían fracasar⁵. Ante el reverdecer de los modelos neurales, el paradigma simbólico ya no puede dar por sentado su dominio y la Imagen Sintáctica de la Mente inicia su defensa frente al acoso de la concepción de la mente asociada a la estrategia conexionista. Así pues, parece que la mejor manera de entender el impacto del conexionismo en la filosofía de la mente consiste en ver en qué medida la Imagen Sintáctica se ve afectada por el surgimiento de los modelos neurales.

En la sección II, examinaremos el contexto en el que el desarrollo de la Inteligencia Artificial parece resultar relevante para nuestra concepción de lo mental. En este sentido, veremos que, hasta mediados de los ochenta, se reivindica la Imagen Sintáctica de la Mente como el único modo de hacer inteligible la relación entre los contenidos mentales de un organismo y las propiedades físicas de su cuerpo. Dedicaremos, por tanto, las secciones III y IV a exponer los elementos centrales de la Imagen Sintáctica de la Mente, así como a indicar la naturaleza de los argumentos que la apoyan y las dificultades con las que tropieza. En las secciones V y VI se explican brevemente las peculiaridades de los modelos conexionistas y se exploran sus consecuencias para los distintos rasgos de la Imagen Sintáctica, con lo que acabará perfilándose una nueva imagen de lo mental, a saber: la Imagen Conexionista. Esta nueva Imagen nos llevará a revisar, en la sección VII, el concepto clásico de teoría cognitiva, así como a iluminar el debate en torno al holismo de lo mental. En la sección VIII se concluye recapitulando los principales puntos y formulando algunas reservas.

3. Jerry Fodor y Zenon Pylyshyn son los representantes más significativos de la lo que aquí denominaremos «Imagen Sintáctica de la Mente»: cf. Fodor (1975, 1983, 1987, 1990) y Pylyshyn (1984).

4. Cf., en este sentido, Dreyfus (1979), Feldman and Ballard (1982), Searle (1980, 1984, 1990), Dennett (1983, 1984), Churchland y Churchland (1990).

5. Hinton y Anderson (1981) y Rumelhart y McClelland (1986) son dos textos centrales en este punto.

II. FISCALISMO, CONTENIDO MENTAL E INTELIGENCIA ARTIFICIAL

Los programas de investigación en Inteligencia Artificial están, en principio, destinados a construir máquinas que simulen las capacidades cognitivas de los seres humanos. Se entiende, con todo, que el desarrollo de tales programas no sólo incrementará nuestra habilidad tecnológica, sino que permitirá aumentar nuestra comprensión de la estructura de la inteligencia, a la vez que mostrar cómo un conjunto debidamente organizado de materiales puede llegar a pensar. La mejora en el conocimiento de nuestras estructuras mentales favorecería el desarrollo de la psicología, mientras que el análisis de cómo ciertas estructuras físicas pueden tener propiedades mentales nos haría avanzar en la resolución del problema ontológico de la relación entre la mente y el cuerpo. Entendemos, precisamente, que la Inteligencia Artificial alcanzará su mayor relevancia en el ámbito de la filosofía de la mente en la medida en que contribuya a la elucidación del mencionado problema ontológico. De tal esfuerzo se derivan otras consecuencias acerca del contenido mental, de la articulación de las teorías científicas, etc., que se mencionarán en la penúltima sección de este capítulo.

El punto de partida de la Inteligencia Artificial es una comprensión preteórica de nuestras capacidades cognitivas. La habilidad para coordinar nuestra conducta con la de otros individuos descansa en gran medida en nuestra capacidad de adscribir deseos y creencias, pues suponemos que es el contenido de nuestros deseos y creencias lo que causa gran parte de nuestra conducta y, por tanto, lo que define qué comportamientos podemos esperar de nosotros mismos y de los demás. La fiabilidad de tales predicciones (y explicaciones) presupone que el comportamiento de los individuos responde a ciertas generalizaciones que relacionan, por un lado, deseos y creencias y, por otro, conductas. En el ámbito de la ciencia cognitiva, se denomina *psicología popular* a la teoría formada por ese conjunto de generalizaciones, si bien la fuerza explicativa de tales generalizaciones descansa en un compromiso ontológico de carácter mentalista, a saber:

La intuición mentalista (IM): Existen actitudes proposicionales (es decir, creencias, deseos, etc.) que tienen un contenido representacional, y que afectan y guían la conducta de quien las posee en virtud de ese mismo contenido.

En el seno de la Inteligencia Artificial se tiende a asumir IM, pues ¿quién se atrevería a dudar de su validez? ¿De qué otra intuición podríamos estar más seguros? Y, sin embargo, no deja de ser misterioso que la semántica de un estado mental pueda alterar la conducta de un organismo. Un ejemplo tomado de Fred Dretske⁶ ilustra las razones para ser re-

6. Dretske (1988, 79-80).

ticentes. Imaginemos una soprano haciendo ejercicios de canto ante una copa, con el fin de romperla mediante el mero uso de la voz. Y así acontece al entonar nuestra soprano una de sus arias. Cuando el aria alcanzó la frase «No destruyas mi frágil corazón», la copa estalló. Coincidiremos en que el contenido del aria en poco ha afectado a la ruptura de la copa, sólo la altura de la voz parece contar. La cuestión que nos inquieta quizá ya se haya imaginado: ¿No ocurrirá lo mismo con los estados mentales que postula la psicología popular? ¿Cómo puede mover el mundo algo tan etéreo como el contenido, como la semántica? ¿No es, acaso, el contenido de nuestros estados mentales tan irrelevante para el movimiento de nuestro cuerpo como el contenido del aria para el estallido de la copa? ¿No tienen los contenidos mentales los mismos problemas que el alma cartesiana para afectar al (o dejarse afectar por el) cuerpo? El origen de esta sospecha deriva de una convicción que emerge con tanta fuerza como la intuición mentalista, a saber:

La convicción fisicalista (CF): El mundo es un sistema cerrado desde el punto de vista de sus propiedades físicas y, en consecuencia, cualquier hecho del mundo ha de tener una explicación física.

Si todos los fenómenos del mundo tienen una explicación física, cualquier explicación que demos de una conducta en términos mentales será, de algún modo, superflua, epifenoménica. Así, tomemos la siguiente generalización

$$(I) \quad M \rightarrow C$$

donde «M» representa un estado mental o actitud proposicional, y «C» una determinada conducta. La convicción fisicalista obliga a que para cada generalización del tipo (I), haya otra generalización como la siguiente:

$$(II) \quad F \rightarrow C$$

donde «F» constituye una propiedad física. Por tanto, debemos entender que es la propiedad física F y no la propiedad mental M lo que es causalmente responsable de la conducta C. Dicho de otro modo, aunque parezca que sea M lo que causa C, en realidad la causa de C es la propiedad física⁷.

Sin embargo, el argumento que acabamos de presentar afecta no sólo a la relevancia causal de los contenidos mentales, sino a la eficacia causal de todas las propiedades de las ciencias especiales como la química, la biología, la geología, etc. En tal caso, los efectos de la convicción fisicalista serían devastadores y reducirían al epifenomenalismo a

7. Cf. Kim (1989, 1990, 1991).

todas las propiedades no físicas. Había que buscar algún remedio a tan nefasta situación. La solución parecía estribar en la posibilidad de encontrar algún tipo de dependencia entre F y M que fuese lo suficientemente fuerte como para respetar la convicción fisicalista, pero no redujese M a F, es decir, no redujese todas las propiedades a propiedades físicas, pues en tal caso la autonomía de las ciencias especiales quedaría anquilada.

La idea de una teoría funcional contribuyó de manera crucial a la elaboración de esa noción de dependencia. En una teoría funcional los términos adquieren su significado por el rol causal que se le adscribe en la teoría a la entidad que designan. Las propiedades funcionales pueden instanciarse físicamente de varios modos. Objetos físicamente muy dispares pueden tener en común la propiedad funcional de «ser un freno» o «ser una llave de contacto». En consecuencia, la relación de realización es una relación unidireccional, pues no incluye la referencia a condiciones necesarias sino únicamente a condiciones suficientes. Diremos, por tanto, que una teoría funcional está físicamente realizada si pueden especificarse condiciones físicas suficientes de la realización de las propiedades que la componen. Una consecuencia de esta unidireccionalidad es que las propiedades funcionales no se reducen a las propiedades físicas y, en consecuencia, se puede reconocer la autonomía de las ciencias especiales. Sin embargo, la relación de dependencia debería ser lo suficientemente fuerte como para respetar la convicción fisicalista; pues nos proporciona, para cada propiedad funcional causalmente eficaz, un conjunto de propiedades físicas que constituye una condición suficiente de su instancia. Ello garantiza que dos objetos o procesos con las mismas propiedades físicas compartan también sus propiedades funcionales y que todo fenómeno, aunque disponga de una explicación funcional, cuente también con una explicación física.

La otra cara de la noción de realización la constituye la noción de *sobrevenir*⁸. Así, podemos decir tanto que un conjunto de propiedades físicas de un objeto *realiza* una determinada propiedad funcional de ese objeto, como que tal propiedad funcional *sobreviene* a ese conjunto de propiedades físicas: sin alteraciones en la base física no puede haber cambios en las propiedades funcionales que sobrevienen a esa base física. Todo ello conduce, en el seno de la ciencia cognitiva, a una reformulación de la convicción fisicalista que resulta, en principio, compatible con la eficacia causal de las propiedades postuladas por las ciencias especiales. De modo más explícito, la nueva versión de la convicción fisicalista quedaría formulada como sigue:

8. Para una presentación de diferentes concepciones de la relación de sobrevenir, cf. Kim (1984, 1990).

El fisicalismo cognitivo (FCG): Una propiedad tiene relevancia causal si y sólo si:

- (1) es una propiedad postulada por una teoría física o
- (2) forma parte de una teoría funcional físicamente realizada, de tal manera que se puedan especificar las propiedades físicas que realizan (o a las que sobrevienen) las distintas propiedades que, desde esa teoría, se atribuyen al mundo⁹.

Sólo las propiedades que satisfagan este principio podrán formar parte de una explicación científica; en consecuencia, la psicología científica únicamente podrá incorporar los compromisos ontológicos de la psicología popular si se puede probar que los estados mentales que esta última postula concuerdan con FCG. Precisamente, la relevancia filosófica de los distintos modelos de Inteligencia Artificial reside fundamentalmente en su habilidad para mostrar en qué medida FCG (heredero actual de CF) es compatible con IM¹⁰. De hecho, los defensores de la Imagen Sintáctica de la Mente entienden que el mejor argumento en favor de esa visión de la mente es que constituye el único modo conocido de conjugar FCG e IM¹¹.

III. LA IMAGEN SINTÁCTICA DE LA MENTE

Supongamos, pues, que los contenidos mentales se fijan en el seno de una determinada teoría funcional, a saber: la psicología popular. Hemos visto que, según el fisicalismo cognitivo, los estados que se postulan en el ámbito de una teoría funcional sólo son propiedades que alteran el mundo si se realizan físicamente. La noción de realización requiere que se indique cómo ciertas variaciones en las propiedades físicas de un organismo afectan a sus estados mentales. En este sentido, la Imagen Sintáctica de la Mente trata de mostrar cómo, para cada contenido mental M de un organismo O, hay un conjunto de propiedades físicas F de ese or-

9. Algunos textos básicos en la discusión actual acerca de las condiciones que debe satisfacer una propiedad para ser causalmente eficaz, son Fodor (1987, 1990), McLaughlin (1989) LePore y Loewer (1987, 1989), Horgan (1989), Kim (1989, 1990, 1991).

10. Esta es ciertamente una de las cuestiones centrales en ciencia cognitiva. Fodor (1987, 1985) es uno de los más fervientes defensores de la compatibilidad de IM y FCG, aunque Dretske (1981, 1987) y Millikan (1984) también se cuentan entre sus defensores. Dennett (1981, 1987c, 1987d) mantiene una posición más instrumentalista respecto a IM, mientras que P.M. Churchland (1986) y P.S. Churchland (1989) insisten en el conflicto entre IM y FCG. Para un mapa de las posiciones más relevantes respecto a esta cuestión: Fodor (1985), Dennett (1969, 1987a, 1987b, 1991) y también Lyons (1990a, 1990b).

11. Bajo el rótulo «Imagen Sintáctica de la Mente» se incluye la teoría representacional de la mente que apadrina Fodor. Se habla de Imagen Sintáctica para contraponer la teoría de Fodor a una teoría representacional derivada de los modelos conexionistas.

ganismo, tal que se cumple la siguiente relación de realización o (de sobrevenir):

F → M.

Como vemos, este requisito exige que para cada distinción semántica haya una distinción física. Mas ¿cómo podría cumplirse esta condición?

Jerry Fodor considera que Turing y, con él, los modelos simbólicos realizan una aportación clave en este punto, pues muestran cómo *la semántica puede afectar al mundo a través de la sintaxis*. Es decir, el único modo de explicar la eficacia causal de los contenidos mentales es postular que éstos se hallan codificados en una estructura sintáctica. Esta es la Hipótesis del Lenguaje del Pensamiento propuesta por Fodor¹². Ese lenguaje, al igual que los sistemas de lógica formal más sencillos, incluye un conjunto de criterios sintácticos para definir sus elementos primitivos, así como sus reglas de formación y derivación. Dos fórmulas de este lenguaje serán distintas en la medida en que difieran en su sintaxis. Hasta aquí nada nuevo. Lo que la Hipótesis de Lenguaje del Pensamiento añade es que:

Toda distinción semántica en nuestros pensamientos encuentra su reflejo en la sintaxis del lenguaje del pensamiento, es decir, si dos pensamientos difieren en su contenido también difieren en su sintaxis. Ahora bien, si esto es así, ya tenemos todo lo que necesitábamos, pues no hay ningún problema en comprender cómo las distinciones sintácticas pueden intervenir en un proceso causal. Las distinciones sintácticas son distinciones formales y es fácil comprender, por ejemplo, cómo la geometría de una llave condiciona las cerraduras que con ella pueden abrirse.

Por tanto, el camino recorrido por la Imagen Sintáctica de la Mente, asociada a los modelos simbólicos, es el siguiente: Los contenidos mentales afectan al mundo porque

- (1) Están codificados sintácticamente y
- (2) tales distinciones sintácticas quedan realizadas en las propiedades físicas del cerebro.

Según nuestro esquema, tendríamos la siguiente relación de realización (o de sobrevenir):

Estados cerebrales — realizan → Fórmulas sintácticas
 Fórmulas sintácticas = codifican ⇒ Contenidos mentales

12. Cf. Fodor (1975, 1987) y Fodor y MacLaughlin (1990).

con lo cual se habría mostrado cómo los contenidos mentales se realizan físicamente y, por tanto, cómo los estados de la psicología popular se realizan físicamente. De este modo, se puede responder ya a la pregunta acerca de cómo los contenidos mentales alteran el mundo, tienen relevancia causal: Los contenidos mentales alteran (y son alterados por) las propiedades físicas del mundo a través de la sintaxis. Para Fodor, la sintaxis es la nueva glándula pineal.

IV. DIFICULTADES PARA LA IMAGEN SINTÁCTICA DE LA MENTE

A pesar de su atractivo, la Imagen Sintáctica de la Mente no deja de plantear graves problemas. En primer lugar, la Imagen Sintáctica supone que hay principios para determinar el contenido mental que están asociados a cada fórmula sintáctica y, en definitiva, a los distintos estados cerebrales. Sin embargo, se ha argumentado poderosamente en contra de tal posibilidad, pues se considera que hay ciertos rasgos peculiares de los contenidos mentales (y, en general, de la semántica) que impiden tal correlación. En concreto, se suelen especificar dos rasgos: la normatividad y la relacionalidad. Este último hace referencia al hecho de que los contenidos mentales de un organismo no quedan fijados por su estructura física, sino por su relación efectiva con el entorno; hasta el punto de que dos organismos físicamente idénticos podrían tener contenidos mentales diferentes en función de las divergencias en sus respectivas biografías, en los entornos naturales y/o sociales en los que se han desarrollado. Por tanto, difícilmente puede haber un conjunto de propiedades cerebrales que sean condición suficiente de la instanciación de un contenido mental. Por otro lado, la normatividad hace referencia a la distinción entre correcto e incorrecto, a la posibilidad de cometer errores. Es esencial al pensamiento el poderse preguntar por si una inferencia es correcta o incorrecta, o si una creencia representa adecuada o inadecuadamente la realidad. Sin embargo, la noción de error tiene difícil cabida en el ámbito de estudio propio de la física, donde sería absurdo buscar principios normativos. El primer reto para la Imagen Sintáctica consiste, por tanto, en esquivar los escollos de la relacionalidad y la normatividad para mostrar que la secuencia cerebro \rightarrow sintaxis \Rightarrow contenido mental realmente existe¹³.

Con todo, estas no son las únicas dificultades que amenazan la viabilidad de la propuesta sintacticista. Hay indicios para defender que los

13. Para seguir la discusión en torno a la normatividad y relacionalidad de los contenidos mentales, cf. Burge (1979), Dretske (1988), Fodor (1987, 1990), McLaughlin (1989), LePore y Loewer (1987, 1989), Horgan (1989), Kim (1989, 1990, 1991), Pettit y McDowell (1986), Putnam (1975), Woodfield (1982).

modelos simbólicos sobre los que descansa la Imagen Sintáctica difícilmente pueden ser modelos que reflejen el modo en que efectivamente funciona nuestra mente. Los modelos sintácticos pueden simular muchas de nuestras capacidades cognitivas, pero no se ajustan al modo como opera nuestro cerebro. ¿Por qué?

Desde los modelos simbólicos, resulta sencillo explicar la productividad de nuestro pensamiento, es decir, la capacidad de generar un número indefinido de pensamientos a partir de una cantidad limitada de elementos. El carácter composicional de la sintaxis puede dar cuenta de este hecho. No obstante, parece que los modelos simbólicos fracasan a la hora de enfrentarse al problema del marco, es decir, a esa habilidad de los seres humanos (y de otros animales) para seleccionar, en un lapso muy breve de tiempo, los aspectos relevantes de una situación para el desarrollo de la actividad que se desea emprender. Cuando entramos en la cocina por la mañana para preparar el desayuno no necesitamos mucho tiempo para determinar que el color de los azulejos de las paredes no afecta al sabor del café con leche. Ante este problema, los modelos simbólicos tienen dos opciones igualmente insatisfactorias. O bien dotar al ordenador de un marco donde se preseleccionen los aspectos relevantes para cada situación o bien que, antes de cada actuación, el ordenador repase y tenga en cuenta todos los datos de que dispone. Esta última opción resulta inviable, pues impediría que el ordenador actuase en tiempo real, no podría simular la rapidez de nuestras actuaciones, al menos si respeta los límites biológicos en cuanto a la velocidad de las transmisiones cerebrales. Por otro lado, la opción de dotar al ordenador con un esquema nos obliga a pagar el precio de una excesiva rigidez que no concuerda con la flexibilidad de nuestras reacciones ante situaciones inesperadas, distintas de las tipificadas en el esquema. En resumen, los modelos clásicos para resolver el problema del marco o bien son demasiado lentos o bien son demasiado rígidos¹⁴.

El problema del marco está vinculado a otras dificultades ante las que sucumben constantemente los modelos simbólicos como, por ejemplo, dar cuenta de la capacidad de generalización espontánea, del aprendizaje, de la degradación paulatina de nuestra actuación, etc. El desarrollo de los modelos conexionistas a principios de los 80 pareció arrojar alguna luz sobre estas cuestiones. A primera vista, se podía ver que propiedades tales como la generalización espontánea o la selección inmediata de los aspectos relevantes para un determinado curso de acción podían ser propiedades asociadas a modelos conexionistas¹⁵.

14. Cf. Dreyfus (1979), Dennett (1983), Churchland y Churchland (1990).

15. Rumelhart y McClelland (1986, págs. 1-4), Smolensky (1988), Clark (1989, cap. 6), Bechtel y Abrahamsen (1991, cap. 2).

V. EL SURGIMIENTO DE LOS MODELOS CONEXIONISTAS

Si los modelos simbólicos toman el lenguaje y su sintaxis como punto de partida para el análisis y simulación de la estructura del pensamiento, los modelos conexionistas buscan una mayor cercanía con la estructura del cerebro. Existe una gran variedad de modelos conexionistas, por lo que la descripción que sigue difícilmente podría hacer justicia a todos ellos¹⁶. Se trata más bien de destacar algunos rasgos que la mayoría de los modelos neurales comparten y que permiten entender su impacto sobre la Imagen Sintáctica de la Mente.

Los componentes fundamentales de los modelos conexionistas son las unidades y sus interconexiones. Las unidades pueden estar activadas o desactivadas, y pueden estar conectadas entre sí por nexos excitatorios o inhibitorios de distinto valor. El valor de esas conexiones y su carácter (excitatorio o inhibitorio) va cambiando a lo largo del proceso de entrenamiento de la red. El valor de activación de una unidad está en función de la fuerza global de los inputs negativos o positivos procedentes de otras unidades con las que se halla conectada, así como del sesgo que puede ir asociado a cada unidad. La influencia que una unidad A puede tener sobre otra unidad B es directamente proporcional, por tanto, al valor de activación de la unidad A junto al valor y naturaleza del vínculo que las une. Como es de esperar, las unidades están agrupadas en redes. En toda red mínimamente sofisticada cabe distinguir unidades de entrada, unidades ocultas y unidades de salida. Un modelo conexionista puede verse como una red de redes de unidades o pautas de conexión.

Un modo de comprender cómo estos sistemas pueden codificar y transformar representaciones es suponer que cada unidad representa un rasgo del mundo. Por poner un ejemplo muy simple, supongamos que hay tres unidades y cada una de ellas representa uno de los siguientes rasgos: «pelo», «vuela» y «pluma». Una conexión positiva entre «pluma» y «vuela» indica que la red cada vez que se activa «pluma» tenderá a activar «vuela», y viceversa. Una conexión negativa actuará, por el contrario, de manera inhibitoria: la activación de «pelo» inhibe «vuela». De este modo, la activación de una unidad opera como la formulación de una hipótesis acerca de la presencia de cierto rasgo en el mundo, mientras que el modo como están interconectadas las unidades revela en qué medida una hipótesis queda reforzada o amenazada por otras hipótesis.

De acuerdo con esto, no puede decirse que los contenidos proposicionales estén codificados en una determinada unidad de una red neural, sino en una red de unidades. Ciertamente, la misma red puede utilizarse para codificar una gran variedad de contenidos proposicionales. Las

16. Se puede encontrar una caracterización más detallada de los rasgos generales de los modelos conexionistas, así como una taxonomía de los mismos, en Rumelhart y McClelland (1986, cap. 2), Smolensky (1988), Hanson y Burr (1990), Bechtel y Abrahamsen (1991).

redes neurales están dotadas de una regla de aprendizaje que hace depender, por ejemplo, el tipo y valor de la conexión entre dos unidades de la frecuencia con que éstas se activan simultáneamente o, al menos, con un cierto grado de proximidad temporal. Esta regla permite que la red determine cómo están relacionados entre sí diferentes rasgos del mundo. Volviendo a nuestro ejemplo, si «pluma» y «vuela» tienden a activarse simultáneamente, la regla de aprendizaje dictará que se refuerce la conexión excitatoria entre ambas unidades, con lo que la red habrá aprendido que, si un objeto tiene plumas, es probable en un cierto grado que vuele; y algo semejante ocurre con las conexiones inhibitorias. De este modo, vemos que una red conexionista aprende modificando sus pautas de conexiones.

Una propiedad importante de las redes neurales es que no se paralizan si no pueden integrar coherentemente toda la información que reciben. Ocurre con frecuencia que los unidades inputs que se activan en un momento determinado representan rasgos del mundo que no son totalmente compatibles entre sí. Los modelos simbólicos se quedarían inmovilizados ante tal situación, a no ser que tal situación hubiese sido prevista con anterioridad y se hubiese programado una determinada solución. En cambio, los modelos conexionistas dibujan un paisaje de soluciones y seleccionan la que permite satisfacer mejor la información recibida, aunque la solución propuesta obligue a dejar de lado algunos de los datos presentados.

Los modelos conexionistas, por tanto, parecen estar en condiciones de dar cuenta de algunas de las capacidades cognitivas que más se resisten a un tratamiento simbólico. En primer lugar, hemos visto cómo puede dotarse a las redes neurales de reglas de aprendizaje que las lleva a generalizar espontáneamente cuando se detecta una cierta correlación (negativa o positiva) entre la presencia de varios rasgos del mundo. En segundo lugar, el comportamiento de las redes conexionistas, al igual que muchas de nuestras habilidades cognitivas, tiende a degradarse paulatinamente. Ello se debe a que el conjunto de condicionamientos que intervienen en el desarrollo de una tarea actúan de un modo «blando», es decir, cualesquiera de esos condicionamientos pueden violarse sin que la tarea se bloquee, si bien cuanto mayor sea el número y relevancia de los condicionamientos vulnerados mayor será el deterioro de la red en la realización de la tarea.

De este modo, se ha llegado a pensar que los modelos conexionistas podrían estar en condiciones de enfrentarse al problema del marco, sin tener que pagar ya sea el precio de la rigidez o de la lentitud. La dinámica de cancelaciones y activaciones que estas redes generan hacen más inteligible nuestra capacidad para enfrentarnos a las peculiaridades de cada situación. Las conexiones entre los distintos rasgos del mundo de los que tiene noticia el sistema están disponibles en las redes de conexiones, y el

sistema realiza un barrido de las mismas cada vez que desarrolla una tarea, con lo cual se seleccionan automáticamente los rasgos del mundo que pueden ser relevantes en una determinada situación. Este barrido puede realizarse en tiempo real porque, a diferencia de lo que ocurre en los sistemas clásicos, no se trata de ir trayendo al módulo central cada uno de los items de información archivados en la memoria, sino que cuando se activan determinadas unidades de salida, el impulso se propaga a lo largo de la red atendiendo a las conexiones que tales unidades de salida mantienen con el resto de las unidades del sistema, con lo cual puede recorrerse la totalidad del sistema en un número bastante reducido de pasos. De este modo, el sistema responde a los imperativos de cada situación sin atenerse a un esquema rígido, sino tomando en consideración toda la información sobre el mundo de que dispone; y, en segundo lugar, puede alcanzar una respuesta en tiempo real porque la dinámica de activaciones y cancelaciones permite recorrer la totalidad del sistema en un número bastante limitado de pasos.

VI. LA IMAGEN CONEXIONISTA DE LA MENTE Y EL PROBLEMA DE LA COMPOSICIONALIDAD

Parece, pues, que los modelos conexionistas están en mejores condiciones que los modelos simbólicos para enfrentarse al problema del marco, para dar cuenta de la capacidad de generalización espontánea, del aprendizaje, etc. Mas, si esto fuese realmente así, los modelos conexionistas estarían ofreciéndonos un análisis alternativo de cómo los contenidos mentales satisfacen FCG. A primera vista, sólo se hace necesaria una ligera modificación respecto a la solución propuesta por la Imagen Sintáctica de la Mente. En los modelos conexionistas los contenidos mentales no se codifican ya en fórmulas sintácticas, sino en redes de actividad. Por tanto, si nuestra mente fuese una red conexionista, entonces parece que el modo como se instancian los contenidos mentales no podría ser, como propone la Imagen Sintáctica, cerebro \rightarrow sintaxis \Rightarrow semántica, sino

redes cerebrales — realizan \rightarrow redes conexionistas
redes conexionistas = codifican \Rightarrow contenidos mentales.

Si esta solución fuese adecuada, la Imagen Conexionista de la Mente aparecería como una alternativa ante la Imagen Sintáctica, pues ambas surgirían como propuestas que pretenden hacer compatibles IM y FCG. En tal caso, la Imagen Sintáctica perdería gran parte de su atractivo, ya que el único argumento en su favor tenía la estructura de una inferencia a la mejor explicación. El surgimiento de una explicación alternativa so-

cavaría trivialmente la fuerza de tal inferencia. Además, la Imagen Conexionista no se presenta sólo como una explicación alternativa, sino como una explicación mejor y más fundamentada. Mejor, porque ofrece mejores perspectivas ante problemas como el del marco, ante los cuales los modelos simbólicos aparecen inermes; y más fundamentada, porque los modelos conexionistas no descansan exclusivamente en una inferencia a la mejor explicación, sino que responden a los rasgos fundamentales de la estructura del cerebro. Por tanto, si este diagnóstico se confirmase, la Imagen Conexionista vendría a sustituir a la Imagen Sintáctica.

Mas no está del todo claro que el diagnóstico vaya a confirmarse. De hecho, se insiste en que los modelos conexionistas fallan en un punto crucial: son incapaces de dar cuenta de la composicionalidad, sistematicidad y productividad de nuestro pensamiento y ello, como veremos, afecta también a la capacidad de la Imagen Conexionista de dar cuenta de la eficacia causal de los contenidos mentales. Desde este punto de vista, la Imagen Conexionista no sería ni siquiera una imagen alternativa y el papel que tendría reservado en el desarrollo de la ciencia cognitiva sería mucho más modesto.

Así, Fodor y Pylyshyn (1988) defienden que los modelos conexionistas no son incompatibles con los modelos simbólicos, pues se ubican en niveles explicativos diferentes. Mientras los modelos simbólicos analizan la estructura de la mente desde un punto de vista cognitivo, los estados que postulan los modelos conexionistas no son estados cognitivos, sino que, en todo caso, podrían contar como una posible implementación de los estados descritos por los modelos simbólicos. El argumento de Fodor y Pylyshyn depende de su concepto de «estado cognitivo»; para ellos, el caso prototípico de estado cognitivo es el tipo de estado que postula la psicología popular. Hasta tal punto que, según Fodor y Pylyshyn, un estado postulado por una determinada teoría es un estado cognitivo si y sólo si (1) se corresponde con un estado de la psicología popular y (2) tiene la misma estructura que los estados de la psicología popular. Lo que impide que los estados conexionistas sean estados cognitivos es este último punto. Fodor y Pylyshyn admiten que se pueda establecer una correlación sistemática entre estados conexionistas y estados psicológicos o representacionales, lo que dudan es que estos dos tipos de estados tengan la misma estructura. El razonamiento es bastante sencillo. Los estados de la psicología popular son productivos, sistemáticos y composicionales. Así, por ejemplo, somos capaces de generar (o de comprender) un número indefinido de pensamientos a partir de un número limitado de elementos primitivos. La única manera de explicar la productividad y sistematicidad del pensamiento es asumir que los mismos tienen *partes constituyentes*, de tal forma que el contenido de un pensamiento complejo se determine a partir del contenido de sus partes constituyentes.

La cuestión es, sin embargo, que los estados conexionistas no tienen partes constituyentes y el significado representacional de una red neural no puede determinarse a partir del contenido de sus partes constituyentes. Por tanto, los estados conexionistas no tienen carácter cognitivo, pues no se ajustan a la segunda de las condiciones definitorias de lo que es un estado cognitivo: en consecuencia, los modelos conexionistas no pueden dar cuenta, desde un punto de vista cognitivo, de la productividad, sistematicidad y composicionalidad del pensamiento. Se sigue, pues, que la Imagen Conexionista no es ni siquiera una imagen alternativa de la mente, dado que es incapaz de explicar los rasgos fundamentales de nuestra estructura cognitiva. Los modelos conexionistas, concluyen Fodor y Pylyshyn, ofrecen como mucho un modo alternativo de implementar los modelos simbólicos.

La respuesta no se hizo esperar. Los defensores de la Imagen Conexionista intentaron mostrar que la conclusión de Fodor y Pylyshyn no estaba fundamentada y, para ello, empezaron a desarrollarse modelos neurales cuyos estados constasen de partes constituyentes y, de ese modo, pudiesen dar cuenta de la productividad, sistematicidad y composicionalidad del pensamiento¹⁷. Smolensky (1987, 1991) propuso un modelo basado en el cálculo vectorial que permitía descomponer cualquier estado conexionista complejo en sus partes constituyentes. Se trataba de modelos conexionistas en los que estados como «taza con café» se representan de manera distribuida en función de la activación o desactivación de unidades más simples tales como «líquido caliente», «superficie curvada de porcelana», «asa para dedo», etc. Una red de este tipo podría representar estados como «café», «taza», «taza con café» y «taza sin café», según el estado de activación de las distintas unidades. En consecuencia, la representación conexionista de un estado como «taza con café» podría adoptar la forma de un vector $\langle a.1, a.2, a.n \rangle$ en el que constase el estado de activación de cada una de las unidades implicadas. Los componentes de ese vector formarían las partes constituyentes del estado «taza con café». Sin embargo, el modelo ha de complementarse con otro vector donde se definan los distintos roles estructurales que el estado «taza con café» puede jugar en las distintas representaciones. Las partes constituyentes del estado «taza con café» vendrían definidas por una operación del cálculo vectorial con los dos vectores mencionados anteriormente.

Disponemos, así, de un mecanismo que nos permite pasar de un vector complejo a sus partes constituyentes y viceversa. Según Smolensky, es posible determinar las especificaciones de una red neural que realizan un vector complejo, pero no se pueden fijar las especificaciones

17. Algunos de los principales modelos conexionistas son Pollack (1988), Hinton (1988), Smolensky (1987, 1991).

a las que sobrevienen cada una de sus partes constituyentes. Además, dado que hay múltiples maneras de descomponer un vector complejo, la determinación de las partes constituyentes de una representación será siempre relativa a un determinado proceso de descomposición. Smolensky no parece darle excesiva importancia a esta falta de unicidad en la descomposición, ya que estima que ello no conduce a un grado preocupante de indeterminación. No todas las estrategias descompositivas aportan los mismos resultados; algunas resultan más fructíferas que otras a la hora de predecir el comportamiento del sistema. Smolensky nos incita, por tanto, a dejarnos guiar por criterios operacionalistas a la hora de elegir el modo en el que los vectores complejos deban descomponerse.

Fodor y MacLaughlin (1990) manifestaron rápidamente su descontento ante la propuesta de Smolensky. Aun suponiendo que su propuesta de análisis vectorial fuese satisfactoria y que pudiesen descomponerse los estados conexionistas complejos en sus partes constituyentes, ello seguiría siendo insuficiente para dar cuenta de la productividad, sistematicidad y composicionalidad del pensamiento. La razón es que la propuesta de Smolensky es puramente instrumental, ya que, a pesar de ofrecer una manera de descomponer los estados conexionistas complejos en sus partes constituyentes, se ve obligado a reconocer que tales partes constituyentes no juegan ningún papel causal en la dinámica del sistema, pues no se pueden determinar las especificaciones de la red neural que realizan las partes constituyentes de un vector complejo. El análisis de Smolensky incrementa nuestra capacidad de explicar y predecir el comportamiento del sistema, pero no da cuenta del proceder real del mismo.

En consecuencia, el precio que Smolensky tiene que pagar para dar cuenta de la productividad, sistematicidad y composicionalidad del pensamiento parece excesivamente alto. El análisis de Smolensky no permite dar cuenta de cómo los estados mentales pueden satisfacer FCG, pues, si bien se puede mostrar cómo se realizan los estados conexionistas complejos, la propuesta de Smolensky no proporciona ningún criterio para determinar cómo se realizan las partes constituyentes de esos estados complejos. Hemos visto, sin embargo, que una parte importante de las transiciones causales entre estados mentales (como las vinculadas a la productividad y sistematicidad del pensamiento) sólo se puede explicar en virtud de la eficacia causal de sus partes constituyentes. Mas, según FCG, para que una parte constituyente de un estado mental sea causalmente eficaz (dé lugar, por ejemplo, a otro estado mental) es necesario que esté físicamente realizada. Por tanto, en la medida en que la propuesta de Smolensky es incapaz de mostrar cómo se realizan físicamente las partes constituyentes de un estado mental, está reconociendo que su planteamiento no puede mostrar cómo los estados mentales puedan ser causalmente eficaces y, en definitiva, cómo IM casa con FCG. De este

modo, si esta insuficiencia del planteamiento de Smolensky fuese expresión de una característica intrínseca del conexionismo, deberíamos concluir, con Fodor y MacLaughlin, que la Imagen Conexionista no puede sustituir a la Imagen Sintáctica, pues esta última sigue siendo el único modo conocido de dar cuenta de la eficacia causal de los contenidos mentales.

El debate no concluye aquí. Abogados de la causa conexionista, como Tim van Gelder (1990) y Andy Clark (1991), tienden a detectar una confusión crucial en la apelación que se hace a la composicionalidad del pensamiento para defender la Imagen Sintáctica de la Mente. Así, Van Gelder concede que hay un requisito genérico de *composicionalidad funcional* que todo tratamiento de la productividad y sistematicidad del pensamiento debe satisfacer. La composicionalidad funcional exige que exista un mecanismo fiable tanto para producir expresiones complejas a partir de número dado de elementos constituyentes como para descomponer las expresiones complejas en sus partes constituyentes. La satisfacción de este requisito de composicionalidad funcional puede garantizarse de varios modos. El más obvio es el que subyace a la Imagen Sintáctica de la Mente, a saber: una composicionalidad concatenativa. Este tipo de composicionalidad es el que permite reconocer en la fórmula compleja « $((p \rightarrow q) \text{ y } p) \rightarrow q$ » los constituyentes « p », « q », « \rightarrow », e « y ». Es decir, utilizamos los mismos criterios para identificar estas partes constituyentes cuando forman parte de la fórmula compleja « $((p \rightarrow q) \text{ y } p) \rightarrow q$ » que cuando se instancian aisladamente o formando parte de otra fórmula compleja. Tenemos un caso de composicionalidad concatenativa cuando las partes constituyentes se pueden reconocer en una parte de la fórmula compleja.

Todo lo que el argumento de Fodor y MacLaughlin muestra es que el análisis vectorial de Smolensky no satisface los requisitos de la composicionalidad concatenativa, pues en el estado conexionista que realiza una fórmula compleja no se pueden reconocer especificaciones que realicen las partes constituyentes de la fórmula en cuestión. Sin embargo, Van Gelder insiste en la posibilidad de diseñar modos de composicionalidad no-concatenativa, es decir, modos de composicionalidad que satisfagan los requisitos de la composicionalidad funcional sin que se haga necesario que se reconozcan las partes constituyentes en la fórmula compleja. Van Gelder propone como ilustración de composicionalidad no-concatenativa la posibilidad de asociar a cada fórmula de la lógica de primer orden un número de Gödel. Ello nos proporciona una función que nos permite determinar el número de Gödel de cualquier fórmula compleja a partir del número de Gödel de sus partes constituyentes, y a la inversa, según nos pide el requisito de composicionalidad funcional. Y, sin embargo, no se pueden reconocer en el número de Gödel de una fórmula compleja los números de Gödel correspondientes a cada una de sus par-

tes constituyentes, con lo cual tendríamos un modo de composicionalidad no-concatenativa.

Si aplicamos este caso a la propuesta de Smolensky, podemos ver que también su análisis vectorial podría satisfacer, en principio, las exigencias de la composicionalidad funcional de un modo no-concatenativo. En efecto, aunque no se pudiese reconocer la realización física de las partes constituyentes de una fórmula compleja en las especificaciones de la red neural que realiza esta última fórmula, podría definirse una función que nos permitiese pasar de la realización de la fórmula compleja a las especificaciones de las redes neurales que realizan sus partes constituyentes, y viceversa. De este modo, resultaría inteligible que las partes constituyentes de un vector complejo cumplieren un papel causal, pues habría una función que, respondiendo a la dinámica efectiva del sistema, nos llevaría de la realización de las partes constituyentes a la realización del vector complejo, sin que haya ninguna necesidad de que en esta última realización quede un rastro reconocible de los estados conexionistas que son responsables de la realización física de las partes constituyentes.

De este modo, podemos también mostrar cómo el hecho de que los vectores complejos admitan varias descomposiciones no debería resultar en principio problemático. Como Smolensky mismo indica, no todas las descomposiciones son igualmente explicativas, algunas nos permiten predecir con mayor exactitud el comportamiento del sistema que otras. Pues bien, ahora cabría decir que, aunque los vectores complejos puedan descomponerse de varios modos, será una de esas descomposiciones la que responda a la dinámica efectiva del sistema, dado que sólo una de ellas dará cuenta de las transiciones efectivas entre los estados conexionistas que realizan las partes constituyentes y los estados que realizan el vector complejo. Lo que hizo que Smolensky se inclinase por una lectura instrumental de su análisis vectorial fue el confundir el requisito de composicionalidad funcional con las exigencias de la composicionalidad concatenativa.

Si esto fuese así, podríamos concluir que la Imagen Conexionista también puede dar cuenta de la eficacia causal de los contenidos mentales, ya que es capaz de compaginar IM y FCG. En consecuencia, la Imagen Conexionista aparecería como una imagen alternativa y la Imagen Sintáctica vería aniquilado el mejor de los argumentos en su favor, a saber: la inexistencia de una explicación alternativa. La Imagen Conexionista dejaría, con todo, abierta la posibilidad de que nuestra mente respondiese a un modelo híbrido en el que, en un entramado básicamente conexionista, ciertos módulos operasen según criterios simbólicos. En cualquier caso, conviene recordar que estamos ante una cuestión abierta y que estas consideraciones tienen una marcada provisionalidad.

VII. TEORÍAS COGNITIVAS, HOLISMO Y LA IMAGEN CONEXIONISTA

El surgimiento de la Imagen Conexionista no sólo afecta al esfuerzo por armonizar IM y FCG, sino que su efecto también alcanza a otras cuestiones en el ámbito de la filosofía de la mente. En esta sección, nos limitaremos a considerar dos efectos secundarios de particular interés. Así, veremos, en primer lugar, que el surgimiento de la Imagen Conexionista conlleva una visión más abierta y pluridimensional del concepto de teoría cognitiva que la que la Imagen Sintáctica de la Mente se ve obligada a preconizar. En segundo lugar, intentaremos mostrar que el desarrollo de la Imagen Conexionista requiere una reconsideración del debate en torno al holismo de lo mental.

Las teorías cognitivas son teorías funcionales. El hecho de que las teorías funcionales puedan definirse a diferentes niveles de abstracción, de manera que cada propiedad funcional pueda plasmarse físicamente de múltiples modos, hizo pensar que las teorías cognitivas deberían despreocuparse del detalle de la realización física y atender exclusivamente a los procesos que se detectan en el plano cognitivo. Esta opción venía a avalar, por otra parte, los programas de investigación en Inteligencia Artificial cuyos modelos parecían muy alejados de nuestras estructuras cerebrales. De este modo, se insistía en que las diferencias físicas entre seres humanos y ordenadores electrónicos no tenían por qué afectar a la naturaleza de sus capacidades cognitivas. Una expresión de este desentenderse de la realización es la insistencia de Fodor y Pylyshyn (1988) en que, si los estados conexionistas no están definidos en un plano cognitivo, el desarrollo de los modelos conexionistas es irrelevante para la comprensión de nuestra estructura cognitiva.

Los modelos conexionistas nacen, por el contrario, de la intuición de que el modo de realización es también relevante, que una buena teoría cognitiva debería incluir diferentes niveles de descripción¹⁸. De hecho, lo que ocurre a los niveles más bajos de descripción condiciona lo que puede ser verdad en los niveles cognitivos de descripción. Como vimos, uno de los datos que tiende a desacreditar a la Imagen Sintáctica como una representación ajustada de cómo funciona nuestra mente es que sólo puede resolver el problema del marco al precio de violar los límites biológicos en cuanto a la velocidad de transmisión de señales y al número de unidades disponibles. Por el contrario, hemos subrayado que una de las virtudes de los modelos conexionistas es que, a primera vista, podrían permitir un tratamiento del problema del marco en términos biológicamente plausibles. En otras palabras, los límites que se fijan en el plano biológico definen los límites a los que debe ajustarse toda teoría plausible de la estructura cognitiva. Ello hace pensar que una teoría de la

18. Cf. Smolensky y otros (1992), Corbí (1993), Clark (1989, 1990), Cussins (1990).

cognición ha de incluir no sólo considerandos cognitivos, sino también referencias a lo que acontece a niveles más bajos de descripción. Parece, pues, que, frente a lo que se da a entender desde la Imagen Sintáctica, las teorías cognitivas han de articular diversos niveles de descripción, y no limitarse a los niveles cognitivos.

Por otro lado, la discusión en torno a la composicionalidad nos ha llevado a descubrir no sólo que las teorías cognitivas han de incluir varios niveles de descripción, sino que la relación entre tales niveles puede ser mucho más compleja de lo que desde la Imagen Sintáctica se pudiera pensar. Así, la relación entre niveles de descripción no tiene por qué responder a la simplicidad de la composicionalidad concatenativa, de manera que se puedan reconocer en la realización física de un estado cognitivo complejo las especificaciones que realizan físicamente cada una de las partes constituyentes de ese estado cognitivo complejo. La función que nos lleva de la realización física de las partes a la realización física del todo será, en general, bastante más compleja. Por tanto, podemos concluir que el surgimiento de los modelos conexionistas altera doblemente la concepción de la teoría cognitiva implícita en la Imagen Sintáctica pues, por un lado, obliga a incorporar diferentes niveles de descripción y, por otro, subraya la complejidad de las relaciones entre los diferentes niveles de descripción.

En segundo lugar, parece claro que la Imagen Sintáctica de la Mente es incompatible con el holismo del significado¹⁹. Si los contenidos mentales se atuviesen a criterios holistas de significado, difícilmente podría establecerse una correlación estable entre fórmulas sintácticas y contenidos mentales. Los criterios de individuación de las fórmulas sintácticas son composicionales y atomistas: la estructura sintáctica de una fórmula compleja es función de sus constituyentes más simples, y la contribución de estos constituyentes a la estructura compleja permanece estable, independientemente de la estructura compleja en la que se inserte. En cambio, si en la identificación de los contenidos mentales nos atuviésemos a criterios holistas, entonces el significado asociado a una fórmula sintáctica variaría en función de las alteraciones en el resto de los contenidos mentales del sistema, con lo que no podría mantenerse una correlación estable entre las propiedades sintácticas de una fórmula y las discriminaciones semánticas del contenido mental que a la misma se pudiesen asociar. Esta incompatibilidad entre la Imagen Sintáctica y el holismo de lo mental incitó a una lectura atomista de la psicología popular, dejando de lado los aspectos de la misma que vienen a subrayar su comportamiento holista²⁰. Por el contrario, la Imagen Conexionista puede permitirse el lujo de recoger estos elementos holistas marginados

19. Cf. Fodor (1987), Fodor y Lepore (1992).

20. En defensa del holismo mental, cf. Davidson (1980), Dennett (1981, 1987e), Block (1986).

por la Imagen Sintáctica, ya que la manera de codificar y transformar representaciones de las redes conexionistas no puede ser atomista y, por tanto, no hay ninguna razón de principio que excluya la posibilidad de correlacionar sistemáticamente sus estados con los estados mentales que postula una lectura holista de la psicología popular²¹. En este sentido, pensamos que la Imagen Conexionista hace posible la rehabilitación del holismo presente en la psicología popular y que la Imagen Sintáctica se ve forzada a negar.

VIII. RECAPITULACIÓN Y CONCLUSIONES

La Imagen Sintáctica de la Mente, asociada a los modelos simbólicos, se propone como la única estrategia de que disponemos para dar cuenta de cómo IM pueda ser compatible con FCG. El surgimiento de los modelos conexionistas parece haber dado lugar a una estrategia alternativa que, además, encuentra mayor apoyo en la estructura del cerebro y permite enfrentarse a problemas cognitivos ante los que los modelos simbólicos parecen fracasar.

Así las cosas, el debate filosófico en torno a la Imagen Conexionista se centra, como hemos visto, en si esta nueva imagen constituye efectivamente una alternativa, es decir, en si realmente es capaz de mostrar cómo los contenidos mentales satisfacen FCG. Se objeta por parte de Fodor, McLaughlin y Pylyshyn que los modelos conexionistas no pueden dar cuenta, desde un punto de vista cognitivo, de la productividad y sistematicidad del pensamiento porque las redes conexionistas carecen de partes constituyentes, es decir, no son composicionales. Hemos intentado mostrar, a la luz de los comentarios de Van Gelder y Clark, que es necesario distinguir entre composicionalidad concatenativa y no-concatenativa, de modo que, en principio, no habría demasiada dificultad en reconocer la composicionalidad no-concatenativa de las redes conexionistas. De este modo, la Imagen Conexionista seguiría apareciendo como una imagen alternativa, por lo que constituirían una seria amenaza para la Imagen Sintáctica, dado que, en principio, los modelos conexionistas no sólo podrían dar cuenta de la productividad y sistematicidad del pensamiento, sino también de otros rasgos del mismo (como el aprendizaje, la generalización espontánea) ante los que los modelos simbólicos parecen naufragar.

No querríamos acabar, sin embargo, sin indicar que algunas cuestiones centrales en torno al problema mente-cuerpo parecen afectar por igual a los modelos simbólicos y a los modelos conexionistas. Si la Imagen Sintáctica tiene problemas para mostrar cómo la normatividad y la

21. Para una discusión de este punto, cf. Ramsey y otros (1990), Corbí (1993).

relacionalidad mental son compatibles con FCG, lo mismo acontece con la Imagen Conexionista. Ello nos hace pensar que la raíz de tales problemas se encuentra en un lugar más profundo y que, tal vez, las exigencias de FCG vayan más allá de lo que nuestras intuiciones fisicalistas nos dictan, por lo que parecería conveniente revisar las condiciones metafísicas que se imponen para reconocer la eficacia causal de los contenidos mentales²².

BIBLIOGRAFÍA

- Bechtel, W. y Abrahamsen, A. (1991), *Connectionism and the Mind*, Basil Blackwell, Oxford.
- Block, N. (1986), «Advertisement for a Semantics for Psychology», en P. French y otros (comp.), *Midwest Studies in Philosophy*, v. 10, University of Minnesota, Minneapolis, 615-78.
- Boden, M. (comp.) (1990), *The Philosophy of Artificial Intelligence*, OUP, Oxford.
- Burge, T. (1979), «Individualism and the Mental», en P.A. French y otros (comp.), *Midwest Studies in Philosophy*, vol. 4, University of Minnesota, Minneapolis, 73-121.
- Chomsky, N. (1957), *Syntactic Structures*, Mouton, La Haya. V. e.: *Estructuras Sintácticas*, Siglo XXI, México, 1974.
- Chomsky, N. (1959), «Review of Skinner's *Verbal Behavior*»: *Language*, 35, 26-58.
- Chomsky, N. (1968), *Language and Mind*, Harcourt, Brace & World, New York. V. e.: *El Lenguaje y el Entendimiento*, Seix y Barral, Barcelona, 1974.
- Churchland, P. S. (1986), *Neurophilosophy: Toward a Unified Theory of Mind/Brain*, The MIT Press, Cambridge, Mass.
- Churchland, P. M. (1989), *A neurocomputational Perspective: The Nature of Mind and the Structure of Science*, The MIT Press, Cambridge, Mass.
- Churchland, P. S y Churchland, P.S. (1990), «¿Podría pensar una máquina?»: *Investigación y Ciencia*, marzo, 18-24
- Clark, A. (1989), *Microcognition*, The MIT Press, Cambridge, Mass.
- Clark, A. (1990), «Connectionism, Competence, and Explanation», en Boden, 1990, 281-308.
- Clark, A. (1991), «Systematicity, Structured Representations and Cognitive Architecture: A Reply to Fodor and Pylyshyn», en T. Horgan y J. Tienson (comp.), *Connectionism and the Philosophy of Mind*, Kluwer Academic Publishers, Dordrecht, 198-218.
- Corbí, J. (1993), «Classical and Connectionist Models: Levels of Description»: *Synthese*, 95, 141-168.
- Cussins, A. (1990), «The Connectionist Construction of Concepts» en Boden, 1990, 368-440.

22. Queremos dejar constancia de nuestro agradecimiento a Tobies Grimaltos, Carlos Moya, Vicente Sanfélix y Josefa Toribio, cuyos comentarios nos han ayudado a mejorar versiones anteriores de este capítulo.

- Davidson, D. (1980), *Essays on Actions and Events*, Clarendon Press, Oxford.
- Dennett, D. (1969), *Content and Consciousness*, Routledge and Kegan Paul, London.
- Dennett, D. (1981), «Intentional Systems», en *Brainstorms*, The MIT Press, Cambridge, Mass., 3-22.
- Dennett, D. (1983), «Styles of Mental Representation»: *Proceedings of the Aristotelian Society*, 83, 213-26. V. e., *La actitud intencional*, Gedisa, Barcelona, 1991, cap. 6.
- Dennett, D. (1984), «Cognitive wheels: the frame problem of AI», en C. Hookway, *Minds, Machines and Evolution*, Cambridge University Press., Cambridge.
- Dennett, D. (1987a), «Three Kinds of Intentional Psychology», en Dennett, 1987e, 43-68.
- Dennett, D. (1987b), «Reflections: Instrumentalism Reconsidered», en Dennett, 1987e, 69-81.
- Dennett, D. (1987c), «True Believers», en Dennett, 1987e, 13-36.
- Dennett, D. (1987d), «Reflections: Real Patterns, Deeper Facts, and Empty Questions», en Dennett, 1987e, 37-42.
- Dennett, D. (1987e), *The Intentional Stance*, The MIT Press, Cambridge, Mass.
- Dretske, F. (1981), *Knowledge and the Flow of Information*, Basil Blackwell, Oxford.
- Dretske, F. (1988), *Explaining Behavior. Reasons in a World of Causes*, The MIT Press, Cambridge, Mass.
- Dreyfus, H.L. (1979), *What Computers Can't Do: The Limits of Artificial Intelligence*, 2.ª edición, Harper & Row, New York.
- Dreyfus, H. L. y Dreyfus S. E. (1988), «Making a Mind versus modelling the Brain», en M. Boden, 1990, 309-333.
- Feldman, J. A. y Ballard, D. H. (1982), «Connectionist models and their properties»: *Cognitive Science*, 6, 205-54.
- Fodor, J. A. (1975), *The Language of Thought*, Crowell, New York. V. e.: *El lenguaje del pensamiento*, Alianza, Madrid, 1984.
- Fodor, J. A. (1983), *Modularity of Mind*, The MIT Press, Cambridge, Mass. V. e.: *La modularidad de la mente*, Morata, Madrid, 1983.
- Fodor, J. (1985), «Fodor's Guide to Mental Representation»: *Mind*, 94, 79-100.
- Fodor, J. A. (1987), *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, The MIT Press, Cambridge, Mass.
- Fodor, J. A. (1990), *A Theory of Content and Other Essays*, The MIT Press, Cambridge, Mass.
- Fodor, J. y Pylyshyn, Z. W. (1988), «Connectionism and Cognitive Architecture: A Critical Analysis»: *Cognition*, 28, 3-71.
- Fodor, J. y McLaughlin, B. (1990), «Connectionism and the Problem of Systematicity; Why Smolensky's Solution Doesn't Work»: *Cognition*, 35, 183-204.
- Fodor, J. y Lepore, E. (1992), *Holism: A Shopper's Guide*, Basil Blackwell, Oxford.
- Hanson, S. y Burr, D. J. (1990), «What Connectionist Models Learn: Learning and Representation in Connectionist Networks»: *Behavioral and Brain Sciences*, 13, 471-489.

- Hebb, D. O. (1949), *The Organization of Behavior*, John Wiley & Sons, New York.
- Hinton, G. E. (1988), «Representing part-whole hierarchies in connectionist networks»: en *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*, Montreal, Quebec, 48-54.
- Hinton, G. E. y Anderson, J. A. (comp.) (1981), *Parallel Models of Associative Memory*, Lawrence Erlbaum, Hillsdale, New Jersey.
- Horgan, T. (1989), «Mental Quasation», en J. Tomberlin (comp.), *Philosophical Perspectives*, 3. *Philosophy of Mind and Action Theory*, Ridgeview Publishing Company, Atascadero (California), 47-76
- Kim, J. (1984), «Concepts of Supervenience»: *Philosophy and Phenomenological Research*, 65, 153-76.
- Kim, J. (1989), «Mechanism, Purpose, and Explanatory Exclusion», en J. Tomberlin (comp.), *Philosophical Perspectives*, 3. *Philosophy of Mind and Action Theory*, Ridgeview Publishing Company, Atascadero (California), 77-108.
- Kim, J. (1990a), «Supervenience as a Philosophical Concept»: *Metaphilosophy*, 21, 1-27.
- Kim, J. (1990b), «Explanatory Exclusion and the Problem of Mental Causation», en E. Villanueva (comp.), *Information, Semantics, and Epistemology*, Basil Blackwell, Oxford, 35-55.
- Kim, J. (1991), «Dretske on How Reasons Explain Behavior», en McLaughlin (comp.), *Dretske and His Critics*, Basil Blackwell, Oxford, 52-72.
- LePore, E. y Loewer, B. (1987), «Mind Matters»: *The Journal of Philosophy*, 84, 630-642.
- LePore, E. y Loewer, B. (1989), «More on Making Mind Matter»: *Philosophical Topics*, 17, 175-191.
- Lyons, W. (1990a), «Intentionality and Modern Philosophical Psychology, I: The Modern Reduction of Intentionality»: *Philosophical Psychology*, 3, 247-269.
- Lyons, W. (1990b), «Intentionality and Modern Philosophical Psychology, II: The Return to Representation»: *Philosophical Psychology*, 4, 83-102.
- McCulloch, W. S. y Pitts, W. (1943), «A logical calculus of the ideas immanent in nervous activity»: *Bulletin of Mathematical Biophysics*, 5, 115-33.
- McLaughlin, B. P. (1989), «Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical», en J. Tomberlin (comp.), *Philosophical Perspectives*, 3. *Philosophy of Mind and Action Theory*, Ridgeview Publishing Company, Atascadero (California), 109-136.
- Millikan, R. G. (1984), *Language, Thought, and Other Biological Categories*, The MIT Press, Cambridge, Mass.
- Minsky, M. A. y Papert, S. (1969), *Perceptrons*, The MIT Press, Cambridge, Mass.
- Neumann, J. von (1956), «Probabilistic logics and the synthesis of reliable organisms from unreliable components», en C. E. Shannon y J. McCarthy (comp.), *Automata Studies*, Princeton University Press, Princeton, New Jersey.
- Pettit, P. y McDowell, J. (comp.) (1986), *Subject, Thought and Context*, OUP, Oxford.
- Pollack, J. (1988), «Recursive auto-associative memory: Devising compositional

- distributed representations», en *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*, Montreal, Quebec.
- Putnam, H. (1975), «The Meaning of 'Meaning'», en H. Putnam, *Philosophical Papers, 2: Mind, Language and Reality*, Cambridge University Press, Cambridge, 215-271.
- Pylshyn, Z. W. (1984), *Computation and Cognition: Toward a Foundation for Cognitive Science*, The MIT Press, Cambridge, Mass. V. e.: *Computación y Cognición*, Debate, Madrid, 1988.
- Ramsey, W., Stich, S. y Garon, J. (1990), «Connectionism, Eliminativism, and the Future of Folk Psychology», en J. Tomberlin (comp.), *Philosophical Perspectives, 4*, Ridgeview, Atascadero, California, 499-533.
- Rosenblatt, F. (1959), «Two theorems of separability in the perceptron»: *Mechanisation of Thought Processes: Proceedings of a Symposium Held at the National Physical Laboratory*, November, 1958, vol. 1, HMSO, London, 421-56.
- Rosenblatt, F. (1962), *The Principles of Neurodynamics*, Spartan, New York.
- Rumelhart, D. E., McClelland, J. L. y el PDP Research Group (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, The MIT Press, Cambridge, Mass. V. e., parcial: *Introducción al Procesamiento Distribuido en Paralelo*, Alianza, Madrid, 1992.
- Searle, J. (1980), «Minds, Brains, and Programs»: *Behavioral and Brain Sciences*, vol. 3, n.º 3, 417-458.
- Searle, J. (1984), *Minds, Brains and Science*, Harvard University Press, Cambridge, Mass. V. e.: *Mentes, cerebros y ciencia*, Cátedra, Madrid, 1985.
- Searle, J. (1990), «¿Es la mente un programa informático?»: *Investigación y Ciencia*, marzo, 10-16
- Selfridge, O. G. (1959), «Pandemonium: A paradigm for learning», *Symposium on the Mechanization of Thought Processes*, HMSO, London.
- Smolensky, P. (1987), «The Constituent Structure of Mental States: A Replay to Fodor and Pylshyn»: *Southern Journal of Philosophy*, 26, 137-160.
- Smolensky, P. (1988), «On the Proper Treatment of Connectionism»: *Behavioral and Brain Sciences*, 14, 1-74.
- Smolensky, P. (1991), «Connectionism, Constituency and the Language of Thought», en B. Loewer y G. Rey (comp.), *Meaning in Mind: Fodor and his Critics*, Basil Blackwell, Oxford.
- Smolensky, P., Legendre, G. y Miyata Y. (1992), «Principles for an Integrated Connectionist/Symbolic Theory of Higher Cognition»: *Informe CU-CS-600-92*, Computer Science Department, University of Colorado, Boulder.
- Van Gelder, T. (1990), «Compositionality: A Connectionist Variation on a Classical Theme»: *Cognitive Science*, 14, 355-384.
- Woodfield, A. (comp.) (1982), *Thought and Object. Essays on Intentionality*, Clarendon Press, Oxford.

ISBN 84-8164-046-8 /



9 788481 640465

La mente humana

Edición de
Fernando Broncano

Editorial Trotta
Superior de Investigaciones Científicas

CONTENIDO

Presentación: <i>Fernando Broncano</i>	11
La tesis de la identidad mente-cuerpo: <i>Eduardo Rabossi</i>	17
El funcionalismo: <i>Manuel García-Carpintero</i>	43
La concepción teológica de los estados mentales y de su contenido: <i>Daniel Quesada</i>	77
Teorías de la arquitectura de lo mental: <i>Jesús Ezquerro</i>	97
El conexionismo y su impacto en la filosofía de la mente: <i>Josep Corbí y Josep L. Prades</i>	151
Teorías del contenido mental: <i>Juan José Acero</i>	175
Causalidad y contenido mental: <i>Manuel Liz</i>	207
Eliminativismo y el futuro de la Psicología Popular: <i>Josefa Toribio Mateas</i>	245
Evolución y lenguaje: <i>Antoni Gomila Benejam</i>	273
El control racional de la conducta: <i>Fernando Broncano</i>	301
Percepción: <i>Vicente Sanfélix Vidarte</i>	333
<i>Qualia</i> : propiedades fenomenológicas: <i>Alfonso García Suárez</i>	353
Conciencia: <i>Enrique Villanueva</i>	385
<i>Índice analítico</i>	401
<i>Índice de nombres</i>	407
<i>Nota biográfica de autores</i>	409