

## Filosofía de la mente: Historia, perspectivas, conceptos, personajes y algunos problemas

*Manuel Liz*

---

Arbor CLVIII, 621 (Septiembre), 65-102 pp.

*En este trabajo se ofrece una introducción a esa área de investigación filosófica de la mente. Se abordarán, en primer lugar, algunos aspectos de su evolución histórica como disciplina diferenciada. En segundo lugar, se describirán las principales opciones existentes a la hora de conceptualizar las relaciones de lo mental con el mundo estudiado por las ciencias naturales. Se precisarán también algunos de los conceptos más recurrentes en la actual filosofía de la mente, proporcionándose a continuación una panorámica general de los autores más representativos en este terreno. Por último, se comentan varios grupos importantes de problemas abiertos. La actual filosofía de la mente nos mostraría, sobre todo, una nueva forma de hacer filosofía que pone constantemente en práctica la inexistencia de fronteras claramente definidas entre filosofía y ciencia.*

---

La filosofía de las últimas décadas se ha diversificado mucho. De su tronco han surgido nuevas ramas. Y algunas de ellas rápidamente se han poblado de vegetación. Entre las abundantes hojas, surgen frutos de apariencia exótica. Este es el caso de la filosofía de la mente. Nuestro *objetivo* aquí es ofrecer una breve introducción a esa nueva y apasionante área de investigación. Por supuesto, serán inevitables las omisiones, los tratamientos parciales y las simplificaciones. Ningún trabajo de estas características es ajeno a un particular punto de vista. Sin embargo, la intención sincera que lo guía es la utilidad. Esperamos

que pueda ser útil tanto al público ya especializado como a un público más amplio, tal vez simplemente interesado en adquirir cierta información básica.

El *plan* es el siguiente. En la primera sección se indican algunas notas imprescindibles para entender la génesis y evolución histórica de la filosofía de la mente tal como ha llegado hasta nosotros. En la segunda sección se presentan algunas de las principales perspectivas que se han ido desarrollando en su seno. A continuación, se aclaran y precisan varios de sus conceptos más importantes y recurrentes. La siguiente sección está llena de nombres propios. En ella se proporcionará una panorámica general de los autores más representativos, de sus relaciones con otros autores y de sus más relevantes aportaciones. Todas las indicaciones bibliográficas se concentran aquí, y podrán servir como orientación a la hora de iniciar una mayor profundización en aspectos concretos. En la sección quinta y última, se señalan y comentan varios de los problemas que aún siguen abiertos en la actual filosofía de la mente. Cada una de estas breves secciones admitiría una lectura separada. Sin embargo, es el conjunto de todas ellas lo que se presenta como marco de referencia.

### 1. Notas históricas

La PREHISTORIA de la actual filosofía de la mente nos remite a la antigüedad. En particular, a los análisis sobre la mente que pueden rastrearse en los primeros filósofos occidentales. Se podría decir que los análisis filosóficos sobre la mente son tan antiguos como la misma filosofía. Pero no podemos olvidar que seguramente nuestra noción de mente no sea exactamente la misma que, por ejemplo, la de Platón o Aristóteles. Tal vez la noción genérica de *alma* sea la que mejor sirva para encuadrar estos análisis. *La filosofía clásica de la mente es estudio filosófico del alma*. Principalmente en dos sentidos. De una parte, suele entenderse que el alma nos pone en conexión con realidades no materiales. De otra, es lo que sustenta nuestro cuerpo. Así los dos temas estelares de toda la filosofía de la mente durante este período son: el del carácter inmaterial y posible inmortalidad del alma y, simultáneamente, el de cuáles pueden ser las funciones del alma en relación a la organización y control de un cuerpo vivo. Para toda la filosofía clásica de la mente, el alma es algo desgarrado, escindido y en constante tensión. A la vez algo *descarnado* y algo *encarnado*. Desde entonces, este drama teórico se vive con pasión.

Decíamos que toda la filosofía clásica de la mente es estudio filosófico del alma. Los datos de partida proceden principalmente de la introspección. Y la filosofía es, sobre todo, especulación. Al lado de algunas curiosas observaciones de carácter más o menos empírico y cercanas a la biología, la introspección y la especulación son los procedimientos metodológicos fundamentales. Una especulación en general guiada por consideraciones metafísicas y, muchas veces, también teológicas.

Un SEGUNDO PERIODO comienza con el desarrollo de la ciencia moderna y llega hasta finales del siglo XIX. El destino de los estudios filosóficos clásicos sobre el alma fue muy diverso. Algunos de ellos fueron asumidos por la medicina y la fisiología tradicionales. La mayoría, sin embargo, se convirtieron en *psicología filosófica*, en tratados metafísico-teológicos acerca del alma. Éste es el ambiente general hasta el desarrollo y consolidación de la moderna *ciencia natural*. Y con ella, las cosas empiezan a cambiar. La filosofía de la mente sigue consistiendo en un estudio filosófico del alma, pero la filosofía comienza a hacerse de otra forma. Estamos en la modernidad. La metafísica y la teología van cediendo terreno al conocimiento científico y a la epistemología. Y racionalistas y empiristas, cada cual a su modo, intentan hacer justicia a la nueva imagen del mundo ofrecida por la ciencia natural.

En relación a la mente, los temas que ahora pasan a un primer plano también tienen que ver con esa nueva imagen del mundo. Y entre todos ellos, destacan dos: el de cómo entender la formación de nuestros conceptos y la elaboración del conocimiento, en particular del conocimiento científico, y el del puesto que le corresponde a la libertad, a la acción humana y a sus productos en ese nuevo mundo descrito por la ciencia natural. Las dos primeras críticas kantianas (y en parte también la tercera) intentan justamente dar una respuesta a estas dos cuestiones. El cambio de enfoque es indudable en muchos sentidos. No obstante, hay algo que sigue permaneciendo constante. Como en el periodo anterior, también aquí la mente parece situarse entre *dos mundos* muy heterogéneos: el abstracto mundo de las ideas y el mundo al que pertenecen nuestros cuerpos.

Un TERCER PERIODO muy importante estaría marcado por el surgimiento, a finales del siglo XIX, de la psicología como disciplina científica encargada de estudiar la mente a través de procedimientos científicos semejantes a los de cualquier otra ciencia natural. La filosofía de la mente deja ya de apoyarse en la introspección y en la pura especulación y se vincula inmediatamente a la nueva ciencia de la psicología.

La *psicología científica* se origina con la puesta en marcha de los primeros laboratorios de investigación empírica acerca de las conductas reflejas. Y este movimiento cristaliza en el *conductismo* como paradigma hegemónico hasta mediados del siglo XX. Definitivamente, la psicología deja de ser estudio filosófico del alma para convertirse aquí en *estudio científico de ciertos tipos especiales de comportamiento*. Éste debía ser el método y el objeto propio de la psicología. El único discurso objetivamente aceptable sobre nuestra vida mental debía ser científico y, de una u otra forma, tenía que referirse siempre al comportamiento.

No puede entenderse todo este proceso al margen del espectacular desarrollo de nuevas *ciencias especiales* que tiene lugar desde finales del siglo XIX. Especialmente, en el campo de los estudios humanísticos. La psicología científica se aprovecha de este empuje. Y la filosofía de la mente comienza a centrarse en cuestiones metodológicas y conceptuales ligadas, sobre todo, a los supuestos e implicaciones del conductismo. En estos años emerge la *filosofía analítica*. Y esta corriente rápidamente se interesa por problemas relacionados con la mente y, en particular, por el propio conductismo. Incluso surge aquí una peculiar forma de conductismo, el llamado *conductismo lógico*.

Para todo el conductismo, los estados mentales son disposiciones a la conducta. Pero, según se entienda que deban ser descubiertas esas disposiciones, nos encontraremos con un conductismo, digamos, psicológico o con el conductismo lógico. Para el primero, elaborado principalmente por Watson y Skinner, sólo pueden llegar a establecerse a través de la investigación empírica, a través de la constatación de correlaciones entre determinados estímulos y determinadas respuestas conductuales. Para el segundo, defendido sobre todo por Ryle, han de ser descubiertas mediante un análisis conceptual del significado de los conceptos mentales. En cualquier caso, dentro del conductismo no hay lugar alguno para un mundo interior de estados y procesos mentales.

Paralelamente a lo anterior, también se perfilan en este período dos importantes tipos de *reacciones críticas* al conductismo muy diferentes entre sí y sumamente importantes. Dos tipos de reacciones que apenas tenían un contenido claro y definido antes de la aparición de la psicología conductista. Una de ellas insiste en que, de alguna forma, los estados mentales deben ser simplemente *idénticos* a determinados estados neuronales. La otra insiste en la *irreductibilidad* de lo mental a ninguna otra clase de fenómenos como los estudiados por la ciencia natural. Ambas reacciones seguirán vivas incluso mucho después de que el conductismo entre en crisis y sea desplazado por otros enfoques. Es más, aún mantienen su vitalidad.

Un CUARTO PERIODO, a partir de los años cincuenta de nuestro siglo, tiene lugar con la crisis del conductismo y con la aparición de los modelos computacionales clásicos de la mente. Rápidamente este computacionalismo ocupa el lugar del conductismo como corriente dominante. La filosofía de la mente vive con intensidad este período realmente revolucionario. Y ya no sólo se vincula a la psicología. Contribuye decisivamente a configurar ese amplio campo de estudios que actualmente conocemos como ciencias cognitivas.

Merece la pena detenernos un poco más en la situación de la psicología al llegar a la segunda mitad del siglo XX. Como decíamos, la psicología ya se ha consolidado plenamente como disciplina científica y el conductismo, por el momento, es el paradigma dominante. Sin embargo, algo sorprendente ocurre. En la década de los cincuenta, el programa conductista entra en *crisis* y deja paso a nuevos estilos de entender la mente. Sobre todo, debido a las grandes dificultades que encuentran los conductistas a la hora de explicar el aprendizaje del lenguaje. Surgen aquí los primeros modelos computacionales y, en términos más generales, la llamada metáfora de la mente como un computador. En unos pocos años, el *computacionalismo* sustituye completamente al conductismo como paradigma dominante. Desde ahora, el conductismo sólo será un programa residual o, como mucho, algo muy restringido a ciertas parcelas de la teoría del aprendizaje en las que no intervenga de manera decisiva el lenguaje. Ya hemos indicado que el lenguaje fué el talón de Aquiles teórico del conductismo. Pero un cambio tan radical de perspectiva sólo resulta explicable si también tenemos en cuenta otras cosas que están sucediendo simultáneamente.

Es en estos años cuando, por ejemplo, surge la *lingüística chomskiana*. También se han ido acumulando desarrollos importantísimos en las *ciencias formales*, especialmente en el terreno más próximo a la lógica. Surgen aquí teorías profundamente novedosas como la teoría matemática de la información o la teoría de la computabilidad. Pero, sobre todo, se produce un *desarrollo tecnológico* que va a cambiar completamente el mundo en la segunda mitad de nuestro siglo: *los computadores*. Únicamente ahora empezamos a darnos cuenta de la importancia que ha tenido este hecho. Ciertamente, se pueden encontrar intuiciones, digamos entre comillas, «computacionales» en Raimundo Lulio o en Leibniz. Pero tan sólo representan un pálido reflejo de lo que podemos tener hoy día al alcance de la mano.

A partir de los años cincuenta, el computacionalismo desplaza al conductismo en psicología. Con ello, la psicología deja ya de entenderse como una ciencia de ciertos tipos especiales de comportamiento para

convertirse propiamente en *ciencia de la mente*, de una mente entendida ahora en términos computacionales. Subyaciendo al comportamiento, la *metáfora de la mente como un computador* permite recuperar los estados y procesos mentales internos a los sujetos, y considerarlos en estrecha analogía con los estados y procesos computacionales que tienen lugar en un computador. Los estados mentales se convierten en estados simbólicos con cierta capacidad representacional, y los procesos mentales en transformaciones realizadas sobre esas representaciones simbólicas. Pero la metáfora de la mente como un computador tiene otras implicaciones si cabe aún más profundas. Según veíamos antes, aunque las reflexiones sobre la mente tengan un largo pasado, la psicología científica y la filosofía de la mente asociada a ella apenas tienen algo más de un siglo de existencia. Pero, todavía una vida más corta cabe asignar a las *ciencias cognitivas*.

La misma expresión «ciencias cognitivas» es muy reciente. El término se acuña a partir de los años 50 coincidiendo con el surgimiento del computacionalismo clásico. En un sentido muy *lato*, las ciencias cognitivas incluirían importantes apartados de todas las disciplinas que, de una u otra forma, estudian la mente. Disciplinas como la lingüística, la lógica, la psicología, la neurofisiología, la inteligencia artificial, la filosofía de la mente, etc., incluso la antropología, la biología evolutiva y la etología. En un sentido más *específico*, donde en lugar de hablar de «ciencias cognitivas» tal vez deberíamos decir «ciencia cognitiva», en singular, se aludiría al complejo campo de estudios que se supone está actualmente emergiendo, con características propias y diferenciadas, a partir de todas las disciplinas anteriores. Aquí es donde se insinúa la idea de que tal vez la mente *no* sea propiedad exclusiva de la psicología. Con las ciencias cognitivas, y mucho más cuando se prefiere hablar de «ciencia cognitiva», en singular, surge el sueño de una ciencia general y unificada de la mente (de nuestra mente y de la mente de los animales, y de los robots, ..., y de los marcianos). Por supuesto, todos los «fans» de las ciencias cognitivas, en su sentido lato o específico, y más estos últimos, confían en que se trate de algo más que de un bonito nombre.

La filosofía de la mente en estos años sufre cambios sustanciales. Dentro de ella, la tradición analítica ha dado lugar a nuevas cosas. La *filosofía de la ciencia* y la *filosofía del lenguaje* se han constituido ya en dos campos con características, estilos y problemas muy específicos. Y gran parte de estas características, estilos y problemas son heredados por la filosofía de la mente. El tema de la reducción de teorías, por ejemplo, se concreta de manera muy particular en el problema de las relaciones entre las concepciones ordinarias y de sentido común sobre

la mente, la llamada *psicología natural* («folk psychology»), las teorías computacionales y otras teorías supuestamente más básicas procedentes del estudio fisiológico del cerebro. Por otro lado, y dado el carácter simbólico, incluso lingüístico, que el computacionalismo clásico asume respecto a las representaciones mentales, hasta el punto de postular en algunos casos la existencia de una especie de *lenguaje del pensamiento*, es comprensible la tremenda incidencia que también la filosofía del lenguaje ha tenido en toda la filosofía de la mente de estos años. Todo esto contribuye a una especialización de la propia filosofía de la mente. Es durante este período cuando se elaboran los principales conceptos y categorías que seguimos hoy día empleando en casi todas las discusiones en este campo.

Entre los problemas ontológicos nucleares del computacionalismo se encuentra el del lugar natural de las propias propiedades computacionales. Y la respuesta más común es que las propiedades computacionales son una clase de propiedades funcionales. Otros temas sumamente debatidos son el del papel de la psicología natural en una ciencia computacional de la mente, el de la necesidad de postular o no un lenguaje del pensamiento, un sistema representacional interno a los sujetos y en gran parte innato, el de cómo las representaciones mentales alcanzan a tener cierto contenido semántico, el de la autonomía metodológica de la psicología como ciencia especial, el del lugar teórico que pueden tener cosas como el desarrollo evolutivo en los sistemas cognitivos orgánicos o las relaciones que mantienen los sujetos con el entorno a través de la acción, el de cómo solucionar las frecuentes explosiones computacionales que se producen en un sistema cuando se quiere prescindir completamente de cualquier resto de actividad interpretativa (también llamado «frame problem» más en el contexto de la inteligencia artificial), etc., etc.

Son, como se ve, muchas las cuestiones abiertas por el computacionalismo. Únicamente hemos mencionado algunas. Por otro lado, también existe un grupo particular de fenómenos mentales que desde siempre han resistido un tratamiento computacional satisfactorio. Se trata de todos aquellos fenómenos mentales en los que intervienen, de una u otra forma, las sensaciones, las imágenes mentales y, sobre todo, la conciencia. En la medida en que consideremos a la conciencia como el rasgo más importante y decisivo de nuestra vida mental, todas las ideas, hipótesis y resultados del computacionalismo pueden llegar a parecer completamente superficiales. Realmente es mucho como para tirarlo todo por la borda. Sin embargo, algunos sostienen que la conciencia, una conciencia no reducible a computación, aún es más.

A partir de los años ochenta, se inicia un QUINTO PERÍODO. El panorama del computacionalismo se enrarece y complica. Aparecen en escena nuevos modelos computacionales para la mente que suponen una ruptura con el computacionalismo clásico, o al menos una profunda revisión crítica. Son los *modelos conexionistas*. El conexionismo pone en cuestión el representacionalismo de tono simbólico y lingüístico asumido por el computacionalismo clásico. Ya no es tan fácil entender los estados mentales como estados representacionales internos a los sujetos y con características simbólicas. Ni los procesos mentales como procesos computacionales consistentes en transformaciones realizadas sobre esos símbolos mentales. La filosofía de la mente se ve obligada a reconsiderar en profundidad todo el computacionalismo clásico y a recuperar viejas ideas e intuiciones de tipo, por ejemplo, neurofisiológico, evolucionista o teleológico. Muchos críticos del computacionalismo clásico también se suman ahora a este amplio movimiento conexionista. Ello hace que la filosofía actual de la mente parezca un inmenso lío.

No hay que olvidar, sin embargo, que el conexionismo es una variedad de computacionalismo. Nuestra mente funcionaría también aquí como un computador. Ciertamente no como un computador clásico (digital, simbólico, con una estructura tipo Von Neumann, etc.), pero al fin y al cabo como un computador (como un computador analógico, no-simbólico, con una estructura de red neuronal, etc.). A finales de los años cincuenta, en los orígenes del computacionalismo clásico y de la inteligencia artificial, se propusieron algunos modelos de este tipo que pronto fueron abandonados por las severas críticas que recibieron por parte de algunos autores de gran influencia en este campo. La insatisfacción con los modelos computacionales clásicos ha acabado reviviendo aquellas viejas ideas. Los actuales modelos conexionistas son más complejos y elaborados que aquellos primeros modelos, y sus éxitos parecen estar haciendo realidad muchas de las antiguas promesas. Pese a todo, parece haber cosas en el computacionalismo clásico, incluso en el mismo conductismo, que merecen ser conservadas. Y esto es hoy día objeto de intensos debates.

Miremos atrás. Hemos estado considerando la prehistoria de la filosofía de la mente como parte de su propia historia. Indudablemente, caben otras opciones. Sin embargo, ver así las cosas permite entender mejor la persistencia de ciertos problemas generales. Hemos distinguido en esa historia cinco períodos fundamentales. Para la filosofía clásica de la mente, la filosofía era especulación y la mente alma. Y el alma debía cumplir un doble papel: tenía que ponernos en conexión con

algo, digamos, inmaterial y abstracto a la vez que debía responsabilizarse de la organización y control de un cuerpo vivo. Tras el impacto de la ciencia moderna, esto se concreta. Ese algo inmaterial y abstracto va a ser en primer lugar el conocimiento. Y todos los aspectos relativos a la organización y control del cuerpo van a pasar siempre por las acciones. Los dos ámbitos de la realidad donde parece inevitable la intervención de un alma, ya mucho más cerca de nuestra moderna mente, son la elaboración del conocimiento y la producción de acciones. Poco después, el conductismo, el computacionalismo clásico y el conexionismo también intentarán mostrar cómo son posibles el conocimiento y la acción. Pero esta vez lo harán ya postulando mecanismos científicamente plausibles para todas aquellas cosas que nuestra mente consigue hacer en estas dos esferas del conocimiento y la acción. Para el conductismo (psicológico), se tratará de mecanismos de canalización de disposiciones conductuales. Y tanto para el computacionalismo clásico como para el conexionista serán mecanismos computacionales, aunque de distinto signo. Las cosas oscilan así entre unos mecanismos capaces de modular las disposiciones existentes en nuestros cuerpos a reaccionar de cierto modo ante determinados estímulos y unos mecanismos altamente abstractos de un género parecido a los mecanismos ejemplificados por nuestros computadores. Más allá de esto, nos hemos encontrado con dos tipos de posturas críticas que no sitúan la mente en algún lugar entre lo corpóreo y lo abstracto. De acuerdo a una de estas posturas, los estados y procesos mentales son algo tan poco subjetivo y tan material como ciertos estados y procesos específicamente localizados en nuestro sistema nervioso central. De acuerdo a la otra postura, lo mental es completamente irreducible a nada que pueda ser estudiado de manera científica. Lo mental forma, por decirlo así, un reino propio y autónomo. Un reino diferenciado esencialmente tanto de la inmaterialidad no subjetiva de las ideas, los programas de computador y otras entidades abstractas semejantes como de la subjetividad tremendamente material de nuestros propios cuerpos vivos, inmersos en un entorno y reaccionando a él de manera unitaria. Y muy diferenciado también de la materialidad no subjetiva de nuestros cerebros.

A través de todo nuestro recorrido hemos detectado, pues, cuatro importantes extremos para lo mental: lo inmaterial pero no subjetivo, lo subjetivo pero material, lo material y no subjetivo y, finalmente, lo subjetivo e inmaterial. El caso es que entre esas cuatro paredes es donde actualmente nos seguimos encontrando.

## 2. Perspectivas

Las perspectivas requieren siempre adoptar un punto de vista determinado. Y el punto de vista que adoptaremos aquí será muy *filosófico*. En esta sección, intentaremos recoger varias de las principales opciones existentes en filosofía de la mente a la hora de conceptualizar las relaciones mantenidas entre lo mental y ese otro mundo que no parece para nada mental y que es estudiado por todas nuestras ciencias naturales, siempre en mayor o menor sintonía con lo físico.

El DUALISMO colocaría a lo mental y a lo físico en dos planos ontológicos radicalmente distintos. Habría tres variedades importantes de dualismo: dualismo de entidades, dualismo de propiedades y dualismo de enunciados. El primero sería más radical que el segundo y el segundo más que el tercero. En el dualismo de entidades, aquello capaz de tener propiedades mentales tiene una realidad completamente distinta de aquello que puede tener propiedades físicas. Son, por decirlo así, dos clases de sustancias muy diferentes. En el dualismo de propiedades, siempre es una misma realidad la que tiene propiedades mentales y físicas. Lo radicalmente distinto aquí son las propiedades, no aquello que es capaz de tener esas propiedades. Por último, el dualismo de enunciados podría admitir un único plano de realidad física tanto para las entidades como para las propiedades insistiendo, sin embargo, en la imposibilidad de traducción, irreductibilidad y autonomía de los enunciados mentalistas respecto a enunciados formulados en términos fisicalistas. La disparidad es aquí sólo una disparidad de lenguajes. Por supuesto, quien quiera más matices, quien, por ejemplo, quiera distinguir entre lo físico, lo material y lo natural, se verá obligado a reformular las anteriores variedades del dualismo.

Enfrentado radicalmente al dualismo, el ELIMINATIVISMO consistiría en la tesis de que no existe realmente nada, o casi nada, de lo que la psicología natural dice acerca de nuestra vida mental. La psicología natural sería una falsa, superficial y mala teoría acerca de nuestros estados y procesos mentales. La psicología científica debería abandonar el rumbo marcado por la psicología natural. Y tal vez también nosotros mismos, en nuestra vida cotidiana, deberíamos dejar de hacer uso del marco conceptual ofrecido por la psicología natural. Hay muchas variedades de eliminativismo. Los eliminativismos computacionales y los eliminativismos neurofisiológicos serían, quizás, los más importantes. Pero no hay que olvidar que el conductismo también admitía lecturas eliminativistas.

Las TEORÍAS DE LA IDENTIDAD son otra manera de ver las relaciones entre lo mental y lo físico. Hay dos grandes variedades de

teorías de la identidad: las teorías de la identidad de tipos o de propiedades (en inglés suele hablarse de «type-type identity») y las teorías de la identidad de casos (en inglés, «token-token identity»). De acuerdo a las primeras, para cualquier propiedad mental existe siempre una propiedad física tal que todo aquello que tenga esa propiedad mental ha de tener esta propiedad física y viceversa. Y, por este camino, toda propiedad mental podría acabar siendo identificable con alguna propiedad física. De acuerdo a las segundas teorías de la identidad, no es necesario que exista tal identidad entre propiedades. Sin embargo, todo aquello que pueda tener una propiedad mental debe ser siempre algo físico, algo capaz de tener propiedades físicas que lo identifiquen. Las cosas que tienen propiedades mentales son siempre cosas físicas. Esta es la identidad que existe.

La identidad de tipos implicaría necesariamente una identidad de casos, pero no a la inversa. Y esto es visto por muchos autores como algo importante, como algo que hace posible cierta clase de fisicalismo no reduccionista. Tanto la mayoría de los funcionalismos como el monismo anómalo de Davidson mantendrían teorías de la identidad de casos, pero no de propiedades.

Para el FUNCIONALISMO, los estados mentales serían identificables con ciertas propiedades funcionales. Propiedades que forman una determinada estructura y son tenidas por entidades físicas que, en principio, pueden pertenecer a clases físicas sumamente heterogéneas. La extensión de una propiedad mental, es decir el conjunto de aquellas cosas capaces de tenerla, surgiría de una reclasificación de las extensiones de muy diversas propiedades físicas. Estados muy diversos de sistemas sumamente heterogéneos desde un punto de vista físico podrían ejemplificar, por ejemplo, la creencia de que Margarita es morena. Y no tendría porqué haber ninguna propiedad física capaz de enmarcar todos esos estados, capaz de identificarse con la propiedad de tener esa creencia. Los estados mentales serían, en otras palabras, múltiplemente realizables. La propiedad, por ejemplo, de creer que Margarita es morena se relacionaría, sin embargo, de manera muy especial con la propiedad de creer que Margarita no es morena. Parece que un mismo estado de cosas no podría ejemplificar o tener a la vez ambas propiedades. Las relaciones de unas propiedades funcionales con otras dan lugar a una estructura funcional. Para el funcionalismo, la psicología debería ir desentrañando las estructuras funcionales que constituyen la mente.

Es habitual entender el computacionalismo como una variedad de funcionalismo. Sin embargo, por sí mismo, el computacionalismo no

estaría obligado a ser funcionalista. El computacionalismo sólo es una tesis acerca de cómo funciona la mente, no acerca de si ese funcionamiento es describible o no en términos puramente físicos. Y el funcionalismo, por su parte, sólo daría una respuesta a la última cuestión. Una respuesta negativa. Para el funcionalismo, el funcionamiento de la mente, ya sea o no un funcionamiento computacional, es únicamente describible en términos funcionales, no físicos. De cualquier modo, entre los temas ontológicos nucleares del computacionalismo se encuentra el del lugar natural de las propias propiedades computacionales.

Gran parte de los funcionalismos son funcionalismos computacionales, clásicos o conexionistas. Pero no todos. Hay también, por ejemplo, funcionalismos directamente basados en la psicología natural. Y funcionalismos humunculares y teleológicos. Los funcionalismos humunculares postularían ciertas subunidades psicológicas capaces de decisión y control de tipo intencional, unidades momentáneamente no analizables. Una de las direcciones del avance teórico consistiría en la progresiva reducción funcional de esos «humunculi». Los funcionalismos teleológicos permitirían la intervención teórica y explicativa de las finalidades, de las funciones que serían adecuadas para cierto sistema en relación a su entorno e historia particulares.

El MONISMO ANÓMALO de Davidson, por su parte, sería un monismo fisicalista de un tipo muy especial. Como monismo fisicalista, sostiene que todo evento mental es un evento físico. Pero es anómalo en el sentido de que rechaza la existencia de leyes estrictas que conecten las propiedades mentales con las propiedades físicas, o las propiedades mentales entre sí. Admite, por tanto una identidad de casos, pero no de tipos o propiedades. No existen estrictas leyes psicológicas ni psico-físicas. Como mucho, la relación que vincula las propiedades mentales con las propiedades físicas es una débil relación de sobrevenida. En la psicología, lo que cuenta es la interpretación de ciertos eventos desde el punto de vista normativo de su racionalidad.

Al lado de las anteriores perspectivas, la noción de SOBREVENIENCIA o superveniencia (en inglés, «supervenience») ha supuesto, en la filosofía contemporánea de la mente, una nueva conceptualización de las relaciones generales que pueden mantener lo mental y lo físico. Se presenta como alternativa frente a los diversos intentos por entender esas relaciones como identidades definicionales entre los conceptos mentalistas y ciertos conceptos físicos (al estilo, por ejemplo, del conductismo lógico), como identidades de tipos entre las propiedades mentales y las propiedades físicas (al estilo, por ejemplo, de los reduccionismos fisicalistas propuestos muchas veces desde el punto de vista de la

neurofisiología), o incluso como identidades sólo de casos (al estilo, por ejemplo, de la mayoría de los funcionalismos y del monismo anómalo de Davidson). La noción de sobrevenida pretende, a la vez, satisfacer todas las exigencias ontológicas del fisicalismo respecto a la determinación de cualquier realidad por la realidad física y dejar abierta la posibilidad de que acaso no sea posible ninguna identidad o reducción de lo mental a lo físico. Ni siquiera una simple identidad de casos. La idea fundamental es la siguiente: si lo mental sobreviene a lo físico, entonces, siendo como de hecho es lo físico, lo mental no podría haber sido diferente a como es. Si se produce una relación de sobrevenida, cualquier variación o cambio en lo que sobreviene debe entrañar una variación o cambio en la base de la sobrevenida; pero no necesariamente a la inversa.

Debemos, por último, hacer referencia a dos perspectivas metodológicas tan generales como diferentes a la hora de aproximarse a los fenómenos mentales. Son las perspectivas del individualismo y del anti-individualismo. Aunque se trate de perspectivas metodológicas, inciden muy directamente en el carácter más o menos local de las relaciones que cabe esperar que mantengan lo mental y lo físico.

El INDIVIDUALISMO consiste en no considerar como psicológicamente relevante nada que no esté exclusivamente determinado por el estado actual de los sujetos. Con ello, desaparecerían de la psicología todas las referencias al entorno y la historia de los sujetos psicológicos. Y, en consecuencia, cualquier referencia a propiedades semánticas no determinadas por sus estados actuales. Se supone que todo lo psicológicamente relevante debe estar de alguna forma dentro de los límites actuales del sujeto. En otras palabras, que el entorno y la historia no pueden importar si no dejan su huella en el sujeto. Y que sólo pueden importar mientras se mantenga tal huella. Esto se aplica inmediatamente a las propiedades semánticas que permiten identificar muchos estados mentales. La polémica en torno a la conveniencia o no del individualismo en psicología ha sido y sigue siendo abundante y apasionada. El «solipsismo metodológico» de Fodor, por ejemplo, sería una consecuencia metodológica de una posición individualista, aunque también pueda ser mantenido por razones independientes.

El ANTI-INDIVIDUALISMO se opone al individualismo. A veces, esta misma oposición, o muy similares, se plantean también en términos de internalismo/externalismo o de propiedades intrínsecas/relacionales. El anti-individualismo considera psicológicamente relevantes ciertos fenómenos del entorno o de la historia de los sujetos. En particular, aquellos fenómenos capaces de hacer variar nuestras asignaciones de

contenido semántico sin que varíen las propiedades individuales, internas o intrínsecas de los sujetos psicológicos. La psicología natural suele considerarse (menos por lo que respecta al fenómeno de la opacidad –véase el siguiente apartado–) fuertemente anti-individualista. En contraste con la manera habitual de entender los procesos computacionales. Por expresarlo de una manera muy gráfica empleada a veces por el propio Fodor, desde un punto de vista puramente computacional podría no haber ninguna diferencia entre un computador que suena ser un robot y ese mismo robot. Los «inputs» y «outputs» procesados, y el propio procesamiento, podrían perfectamente ser los mismos.

### 3. Los conceptos

En el apartado histórico, señalábamos que la mayoría de los conceptos actualmente presentes en las discusiones de filosofía de la mente han sido elaborados al hilo del desarrollo del computacionalismo clásico como perspectiva dominante en el campo de las ciencias cognitivas. Tal vez sea hora ya de nuevos conceptos y lenguajes. De todas formas, aún no los tenemos. Y esto obliga a tomar precauciones. En las aclaraciones que a continuación se proponen de algunos de los conceptos más importantes de la filosofía de la mente se intenta tener esto muy en cuenta.

La INTENCIONALIDAD suele considerarse, siguiendo a Brentano, una de las marcas características de lo mental. Son estados mentales intencionales los estados mentales que tratan sobre algo, que se orientan inevitablemente hacia un objeto. Las creencias y los deseos son típicamente estados mentales intencionales. Pero también lo son estados como recordar, sospechar, imaginar, dudar, etc. Sus objetos intencionales son aquellas cosas que son creídas, deseadas, recordadas, sospechadas, etc. Los objetos intencionales suelen ser descritos a través de oraciones subordinadas a los verbos de creencia, deseo, sospecha, etc. así, por ejemplo, el objeto intencional de la creencia en que Margarita es morena sería la proposición de que Margarita es morena. Ese es su contenido intencional. En otras palabras, el contenido intencional de un estado intencional es un contenido proposicional. Por ello, a los estados mentales intencionales se les llama también «actitudes proposicionales» y los verbos empleados en su descripción son llamados «verbos de actitud proposicional». Pese a todo, sigue siendo un problema pendiente determinar si ésta es la única reconstrucción posible, si los estados mentales intencionales son únicamente describibles como actitudes

proposicionales, como actitudes psicológicas hacía un contenido proposicional. Otro problema constante aquí es el de si es posible explicar en términos naturalistas la intencionalidad de lo mental, si es posible explicar de manera naturalista el hecho de que algunos de los estados internos de un sistema adquieran una intencionalidad genuína y originaria, vinculándose así esencialmente a un determinado contenido proposicional. Y una parte importante del anterior problema, que sólo será realmente grave para quien sea realista respecto a la intencionalidad, lo constituye la búsqueda de otros fenómenos en la naturaleza capaces de admitir descripciones parecidas a las descripciones intencionales que hacemos de los estados mentales intencionales, fenómenos que sin ser propiamente fenómenos mentales puedan, no obstante y gracias a esa similitud, considerarse piezas básicas de lo mental.

No hay que confundir la intencionalidad con la INTENSIONALIDAD. La intencionalidad es una propiedad de lo mental. La intencionalidad es un fenómeno lingüístico característico de aquellas expresiones que describen fenómenos intencionales, pero no sólo de ellas. Podemos encontrarnos con un contexto lingüístico intensional siempre que resulte problemática una cuantificación existencial o la sustitución de equivalentes lógicos. Los contextos modales, los contextos de descripción de una acción, de una ley de la naturaleza, etc., son también contextos intensionales. Las descripciones que hacemos de los estados mentales intencionales no admiten una cuantificación existencial. De «creer que Margarita es morena» no puede simplemente derivarse «existe Margarita». Tampoco admiten la sustitución de equivalentes lógicos. «Creer que Margarita es morena» no es lo mismo que «creer que la persona que hace la mejor Lasaña Renacimiento del mundo es morena», incluso aunque Margarita sea esa persona. Un contexto intensional es, empleando otra terminología, un contexto opaco, no-transparente.

En un sentido muy general, el contenido semántico o, simplemente, el CONTENIDO, de un estado mental sería el conjunto de sus propiedades semánticas. Propiedades tales como su significado, su referencia, sus condiciones de verdad, etc. En un sentido más específico, el contenido de una determinada actitud proposicional estaría constituido por la proposición hacia la cual se mantiene esa determinada actitud. Por ejemplo, el contenido de mi creencia de que Margarita es morena sería la proposición de que Margarita es morena. Una psicología científica que se vea obligada a prescindir, como lo hace el computacionalismo, de las propiedades semánticas de nuestros estados mentales no podrá sin más describirlos y clasificarlos según su contenido.

Como mucho, utilizará descripciones y clasificaciones que aludan a los posibles contenidos que sean capaces de tener nuestros estados mentales en relación a las propiedades no semánticas que ejemplifiquen. Este es el origen de la tan discutida, aún hoy día, distinción entre contenidos en sentido restringido («broad contents» o «wide contents») y contenidos en sentido estrecho («narrow contents»).

Desde el fracaso del conductismo, el interés por los estados y procesos psicológicos internos ha hecho que la noción de representación ocupe un primerísimo plano dentro de las ciencias cognitivas. REPRESENTAR es hacer que cierto estado de cosas lleve información sobre otro estado de cosas. Las representaciones pueden ser convencionales o naturales. Pueden también ser de tipo lingüístico o no lingüístico. Las oraciones que aquí mismo están escritas representan lingüística y convencionalmente algo. Llevan información sobre algo, de carácter por cierto bastante abstracto. La danza de las abejas a la puerta de sus colmenas también representa algo, esta vez de una manera no convencional ni lingüística. Para casi toda la filosofía contemporánea de la mente, los contenidos de nuestros estados mentales intencionales deben ser representaciones. El que esas representaciones puedan ser, en el fondo, naturales, o en cambio tengan siempre una ineliminable carga convencional, y el que se articulen con una estructura similar a la de un lenguaje, o más bien a través de cosas como, por ejemplo, los modelos mentales o incluso las redes conexionistas (en cuyo caso hasta sería un problema si podemos seguir hablando de representación), son cuestiones que dividen y enfrentan todas las posiciones actuales en este campo.

Pero, cambiemos de esfera de lo mental. No sólo hay estados mentales intencionales. También hay estados mentales que no se orientan a nada, que no tienen ningún objeto intencional. Son todos esos estados que se caracterizan por implicar variaciones en el CARÁCTER CUALITATIVO o fenoménico de nuestra experiencia. El carácter cualitativo o fenoménico de algunos estados mentales, de las sensaciones o los dolores por ejemplo, es la peculiar manera como son experimentados por un sujeto. La idea básica que aquí entra en juego es la siguiente: se supone que aunque alguien pudiera conocer todo lo que la ciencia es capaz de llegar a conocer acerca de los dolores de muelas, por ejemplo, sin embargo podría no saber en qué consiste esa experiencia tan especial que es tener un dolor de muelas. Para conocer esto último, no habría más camino que tener dolores de muelas. Lo que se conoce al experimentar un dolor de muelas sería el especial carácter cualitativo o fenomenico que tienen los dolores de muelas. O un dolor de muelas en particular.

Junto con la intencionalidad y el peculiar carácter cualitativo de algunos estados mentales, la CONCIENCIA es otra de las esferas más importantes de lo mental. Ser consciente significa, entre otras cosas, ser capaz de conocer introspectivamente nuestra vida mental. Y significa, además, reconocerla como propia. Esta serie de capacidades, que motivaron el dualismo cartesiano entre la sustancia pensante y la sustancia extensa, y otros muchos dualismos y monismos mentalistas, han sido tradicionalmente refractarias al análisis científico. Con la conciencia entran en escena nuevos elementos pertenecientes a nuestro ser personas, a nuestra identidad personal, a nuestra libertad y responsabilidad moral, etc., que aumentan enormemente la complejidad de lo mental. Sólo muy recientemente se han presentado teorías no dualistas, teorías de inspiración naturalista, específicas sobre la conciencia.

La vida mental de un sujeto condiciona su acción. La ACCIÓN es un tipo de conducta. Y la conducta es un tipo de comportamiento. Los comportamientos podrían ser definidos como la variación o conservación en el tiempo de una determinada magnitud. Las conductas serían comportamientos originados por la dinámica interna de un sistema dotado de cierta especial complejidad estructural. Y las acciones, por último, serían conductas originadas por una dinámica interna de tipo intencional, por un mundo interno de creencias, percepciones, deseos, recuerdos, etc. Las acciones son siempre, en sentido propio, acciones intencionales. En otras palabras, conductas deliberadas. Haciendo estas precisiones terminológicas, podríamos distinguir entre el comportamiento de los sistemas en general, sistemas de tipo mecánico, socio-económico, ecológico, etc., y la conducta que manifiestan, por ejemplo, los sistemas orgánicos individuales. Y estas dos cosas de las acciones que llevan a cabo sistemas como nosotros. El que el origen de las acciones deba radicar en algo interno a los sujetos es cuestionado por el anti-individualismo. Sin embargo, sigue siendo la concepción más generalmente aceptada.

El tema de la acción ha sido analizado desde muchos puntos de vista: desde la pragmática, desde la hermeneútica, desde la filosofía del lenguaje, desde la filosofía de la mente, desde las teorías formales de la racionalidad como la teoría de la decisión, etc. Y los problemas más importantes en este campo, problemas estrechamente interrelacionados y de muy larga tradición, han sido el de la individuación de las acciones, el de la reducción o no de las acciones a conductas, y éstas a su vez a formas más básicas de comportamientos, el de la coordinación entre la justificación de la acción en base a razones, a

ciertos conjuntos de creencias, percepciones, deseos, recuerdos, etc., relacionados de acuerdo a cierta trama racional, y su explicación en términos causales, y el problema de la caracterización misma de la noción de «acción racional», su reducción a algoritmos o heurísticas, los supuestos y aplicaciones de las teorías formales de la racionalidad, las relaciones entre la racionalidad práctica y la racionalidad epistémica, etc.

La RACIONALIDAD puede realmente predicarse de muchas cosas. Pero, sobre todo, cabría distinguir una racionalidad epistémica y una racionalidad práctica. La racionalidad epistémica se referiría a cierta manera como pueden articularse nuestras creencias, individuales o colectivas, dando lugar a cuerpos de creencias justificadas y aptas para convertirse en conocimientos en el caso de resultar verdaderas. La racionalidad práctica se referiría a cierta manera como pueden articularse nuestras creencias, deseos, intereses, fines, etc., individuales o colectivos, de forma que sean capaces de producir acciones justificadas y adecuadas tanto a la realidad como a nuestros ideales. Tal como hemos dispuesto las cosas, la racionalidad práctica englobaría a la racionalidad epistémica. ¡Y es bueno que así sea! El fuerte contenido evaluativo y normativo presente en el concepto de racionalidad hace que cuando se discutan estos temas siempre se acabe planteando, por un lado, lo que es o no deseable tener como racional y, por otro, si de hecho somos o podemos ser tan racionales como sería deseable.

La PSICOLOGÍA NATURAL (llamada también, a veces, psicología de sentido común, psicología de las actitudes proposicionales, psicología intencional, psicología cotidiana, psicología ordinaria o psicología popular, traduciendo literalmente en este último caso la expresión inglesa «folk psychology») sería el marco conceptual en el que habitualmente nos vemos a nosotros mismos como sujetos psicológicos. Como sujetos con una vida mental llena de creencias, deseos, percepciones, recuerdos, etc., llena de estados y procesos mentales en definitiva, como agentes que también llevan a cabo acciones intencionales y como personas con atributos como la conciencia, la racionalidad, la libertad o la responsabilidad moral. La psicología natural se puede rechazar, asumir, analizar, etc., pero estrictamente, y esto es muy importante, no es una corriente de pensamiento dentro de las ciencias cognitivas. No más que la física natural lo es dentro de las ciencias físicas. El problema fundamental en este tema consiste en poner en relación ese marco conceptual con las teorías de la psicología científica y con las teorías, más básicas, acerca de la génesis, estructura y funcionamiento de nuestros cerebros. Las soluciones a este problema varían enormemente desde

aceptar a la psicología natural como la única guía capaz de orientar la elaboración de teorías científicas sobre la mente hasta proponer su abandono y eliminación en favor de otros tratamientos. La gran sensibilidad de la psicología natural a los peculiares entornos e historias de los sistemas que conceptualiza, el papel básico que tienen en ella las consideraciones de racionalidad y su desconocimiento de lo que puede estar ocurriendo dentro de nuestros cuerpos cuando pensamos constituirían algunos de los principales obstáculos para su coordinación con las teorías científicas sobre la mente.

#### 4. Los personajes

Comencemos ya nuestro repaso de personajes. Lo haremos a partir del momento en el que el conductismo se convierte en el primer espejo científico en el que puede por fin mirarse una filosofía de la mente. Entre los más destacados representantes del conductismo en psicología se encuentran B. SKINNER y J. WATSON. Véanse, por ejemplo, Skinner (1933 y 1957) y Watson (1913 y 1930). Tanto el conductismo psicológico como el conductismo lógico intentan reducir lo mental a disposiciones conductuales. Pero el primero, a diferencia del segundo, pretende llegar a esa reducción a través de la investigación empírica, sobre todo de los mecanismos de condicionamiento operante, no sólo mediante análisis conceptuales. Por parte del conductismo lógico, G. RYLE (1949) fue su principal exponente en los momentos de mayor esplendor de la filosofía analítica entendida como análisis del lenguaje. Defendió que el lenguaje mentalista no hablaba, en el fondo, más que de la conducta. Defendió también que esas conexiones entre lo mental y la conducta eran conexiones de significado. Un análisis conceptual de los conceptos mentalistas debía bastar para averiguar qué disposiciones conductuales estaban detrás de ellos. Actualmente, ya nadie piensa que estas reducciones definicionales sean posibles.

Justo en la cresta de la ola conductista, la crítica de N. CHOMSKY (1959) al conductismo de Skinner intentando explicar el aprendizaje del lenguaje fué demoledora. Los trabajos revolucionarios de Chomsky en lingüística son de sobra conocidos. Véase, por ejemplo, Chomsky (1957 y 1965). Sus tesis sobre el innatismo de ciertas estructuras lingüísticas, a pesar de haber sido tan influyentes, por ejemplo en Fodor, son ya otra cosa. Véase Chomsky (1980 y 1985).

También en la década de los cincuenta se sitúa la obra pionera de A. TURING en el campo de la inteligencia artificial. El concepto

de máquina de Turing sirvió para aclarar enormemente la noción de computabilidad y decidibilidad. Su experimento mental llamado «el juego de imitación», véase Turing (1950), intenta persuadirnos de que manifestar una conducta lingüística inteligente resulta decisivo a la hora de atribuir propiedades mentales. Paradójicamente, esto contribuyó a poner en un primerísimo plano los procesos internos. Si una máquina construida por nosotros manifiesta comportamientos inteligentes, lo hace por el especial programa que ejecuta. El libro de Haugeland (1985) es una de las introducciones filosóficamente más interesantes que actualmente existen al campo de la inteligencia artificial.

Después del ocaso del conductismo, el computacionalismo ocupó rápidamente una posición hegemónica dentro de las ciencias de la mente. Y el funcionalismo se convirtió en la perspectiva filosófica más extendida. Junto con Lewis, Putnam y Fodor, W. SELLARS es uno de sus creadores, tal vez el primero a pesar de no haberla desarrollado con detalle. También es uno de los filósofos que más han insistido en el carácter teórico de la psicología natural. La psicología natural sería una teoría, una teoría empírica y, además, una teoría empírica posiblemente falsa. Véase Sellars (1956). Tanto D. LEWIS como H. PUTNAM, en cambio, sí han hecho propuestas funcionalistas concretas. El primero hace derivar el funcionalismo respecto a lo mental de una interpretación funcional de todos los conceptos teóricos en general. Y no se trata de un funcionalismo necesariamente computacional. Para poder definir funcionalmente los estados mentales, basta con disponer de una red teórica que conecte suficientemente unos estados mentales con otros y con determinados «inputs» y «outputs». Y esto puede perfectamente ofrecerlo la misma psicología natural. El funcionalismo de Lewis, además, no está reñido con una teoría de la identidad de tipos. Entre otros trabajos suyos, Lewis (1972 y 1980) serían puntos de referencia fundamentales. Putnam, por su parte, defendió la hipótesis de que la estructura funcional de la mente era análoga a la de una máquina de Turing. Suele llamarse Turing-funcionalismo a esta variedad de funcionalismo computacional. Véase Putnam (1960, 1965 y 1967). Sin embargo, sus convicciones funcionalistas parecen haber entrado recientemente en crisis. Así como lo funcional es múltiplemente realizable en lo físico, Putnam sostiene ahora que lo mental también es múltiplemente realizable en lo funcional. Véase, por ejemplo, Putnam (1988 y 1992).

Putnam es también el indiscutible inventor de los experimentos mentales acerca de «Tierras Gemelas», «cerebros en cubetas» y otras curiosidades semejantes. En base a ellos, y simultáneamente a T. BURGE, ha presentado algunos de los mejores argumentos contra el in-

dividualismo en psicología. Véase Putnam (1975a y 1975b). Es en este contexto donde también surge la idea de un socio-funcionalismo. Los experimentos mentales y ejemplos de Burge tienen, sin embargo, una ventaja nada desdeñable sobre los de Putnam: son más creíbles. Véase Burge (1979 y 1986).

J. FODOR merece en todo este contexto una consideración muy especial. Su funcionalismo adopta la forma de una teoría representacional y computacional de la mente. El computacionalismo de Fodor ha sido, y tal vez siga siendo, la alternativa más global y elaborada que se ha presentado después del fracaso del conductismo. Los estados mentales son entendidos como relaciones computacionales que un sujeto mantiene con ciertas representaciones, y los procesos mentales como operaciones computacionales realizadas con esas representaciones. El conjunto de representaciones que articulan nuestra vida mental tiene una estructura lingüística. Pensamos a través de un sistema simbólico de representaciones internas. Pensamos en términos de un lenguaje del pensamiento cuyos componentes básicos deben ser, además, innatos. El objetivo de la psicología científica es desentrañar ese lenguaje en el que nuestras neuronas procesan la información, y hacer explícitas las relaciones y operaciones computacionales que allí se producen. Y la psicología científica puede estudiar todo esto sin conocer demasiados detalles sobre el cerebro. Puede hacerlo con la guía segura de la psicología natural. La psicología es una ciencia especial cuyas leyes no son reducibles a las de ninguna otra ciencia más básica. Las propiedades mentales son propiedades funcionales no reducibles o identificables con propiedades físicas. Son propiedades que se identifican por el papel que ciertos elementos desempeñan en la dinámica del sistema cognitivo en su conjunto. Se identifican por sus características sintácticas. Esto compromete inevitablemente con cierto individualismo, con un «solip-sismo metodológico» que coloca fuera del alcance de la psicología científica las relaciones del sujeto con todo lo que sea externo a él. Las propiedades semánticas de las representaciones, por ejemplo, sólo podrán tener importancia psicológica en cuanto tengan algún correlato sintáctico.

Fodor se ha esforzado a lo largo de toda su obra en ser fiel a un lema: aceptar plenamente todas las consecuencias que se sigan de premisas plausibles. Claro que, a veces, aunque tales consecuencias sean perfectamente defendibles una por una, el conjunto puede llegar a tener menos plausibilidad que las mismas premisas. Algo de esto parece estar pasando actualmente con grandes apartados de su programa computacional. Véanse todas las referencias de Fodor que se-

ñalamos en el apartado bibliográfico. Asimismo, véase también, Loewer y Rey, (Eds.) (1991).

Ha habido una gran constelación de autores que, de una u otra forma, han contribuido tanto a desarrollar el programa de Fodor como a analizar sus supuestos e implicaciones. D. MARR (1982) constituye un perfecto ejemplo de investigación funcionalista y computacional clásica en el área de la visión. J. GARCIA-ALBEA (1986) y M. VEGA (1984) son buenas introducciones en castellano al computacionalismo clásico. Y A. RIVIERE (1991) proporcionaría un perfecto contrapunto crítico.

En otra línea, el trabajo de H. FIELD (1978) fue tremendamente clarificador respecto a la semántica que podía esperarse y no podía esperarse dentro de un programa computacional para el estudio de la mente. A él se debe, entre otros como, por ejemplo, G. HARMAN (1973) la idea de una semántica del papel conceptual para el lenguaje del pensamiento. Lenguaje del pensamiento que, tanto para Field como también para Harman, no tiene las características innatas y universales que postula Fodor, sino que es, más o menos, el lenguaje natural de cada uno. Tal semántica del papel conceptual sería autónoma de cualquier teoría del significado basada en las relaciones causales o informacionales con el entorno. La vinculación de esta idea de la posibilidad de una semántica del papel conceptual con la distinción entre un contenido en sentido restringido («narrow content») y un contenido en sentido amplio («broad content» o «wide content») es directa. Respecto a esta última distinción, véase Pettit y McDowell (Eds.) (1986) y Woodfield (Ed.) (1982).

Ha sido también sumamente controvertido el tratamiento que cabe hacer de las imágenes mentales dentro del computacionalismo clásico. Z. PYLYSHYN (1984) ha argumentado con pasión que todas las representaciones mentales tienen una estructura lingüística, incluyendo las procesadas por nuestra imaginación. Las imágenes mentales deben ser tratadas también como sentencias de un lenguaje del pensamiento. Tal defensa apasionada del lenguaje del pensamiento le ha llevado últimamente a criticar, junto con Fodor, muchas de las pretensiones del conexionismo. Véase, Fodor y Pylyshyn (1988). Dejando aparte el conexionismo, otros intentos por hacer respetable la idea de representaciones mentales no lingüísticas o, en otras palabras, la idea de que no pensamos a través de palabras, sentencias y transformaciones de unas sentencias en otras, sino a través de modelos mentales, son P. JOHNSON-LAIRD (1983), J. HOLLAND et al. (1989) y C. MCGINN (1989).

Al hablar del lenguaje, no podemos olvidarnos de D. DAVIDSON. Se podría decir que las aportaciones de Davidson a la filosofía se reparten entre la filosofía del lenguaje, la filosofía de la acción y la filosofía de la mente. Pero esto sería simplificador. En Davidson, el lenguaje, la acción y la mente se encuentran indisolublemente unidos. Tal vez, lo más conocido de su filosofía de la mente sea su posición conocida como monismo anómalo y su defensa del papel de la suposición de una racionalidad máxima a la hora de traducir lenguajes, de interpretar acciones y de atribuir pensamientos. Véase especialmente Davidson (1963, 1970 y 1984) así como LePore y McLaughlin (Eds.) (1985) y LePore (Ed.) (1986). Una buena crítica del requisito de una racionalidad máxima se encuentra en CHERNIAK (1986) y en NISBETT y ROSS (1980).

Dos clásicas ideas de P. GRICE respecto al significado han sido asimismo tremendamente influyentes en la filosofía contemporánea de la mente. Una de ellas consiste en la distinción entre las propiedades semánticas de los lenguajes públicos, de una parte, y las propiedades semánticas de las intenciones de un hablante, de otra. Las primeras se deberían reducir a las segundas, tal vez con la ayuda de un concepto preciso de convención. La otra idea consiste en la presunción de que existe algún tipo básico de significado natural totalmente independiente de nuestras interpretaciones y convenciones. Véase Grice (1957 y 1969). Las dos ideas juntas constituyen un marco conceptual en el que gran parte de la filosofía contemporánea de la mente se encuentra a gusto. Y, de manera muy particular, la hipótesis del lenguaje del pensamiento. Todas las propiedades semánticas deben tener su origen en algo natural, y el origen natural de las propiedades semánticas de nuestros lenguajes públicos se encuentra localizado en nuestros pensamientos. Los análisis críticos de esta concepción por parte de S. SCHIFFER son inteligentes y rotundos. Acaso excesivamente pesimistas, llegando a concluir que no es posible ninguna teoría del significado, y con ello ninguna teoría del contenido mental, que satisfaga simultáneamente todas las condiciones que normalmente se le imponen. Véase Schiffer (1987 y 1991).

Otros autores también han sido mucho más moderados que Fodor en relación al alcance del propio computacionalismo. Las teorías psicológicas, para W. LYCAN, por ejemplo, deben poder postular unidades de control y decisión inteligentes, aunque siempre menos inteligentes que el sujeto psicológico total que se esté estudiando. Esta posición suele conocerse como funcionalismo humuncular. Según la teoría psicológica avance en profundidad explicativa, esas unidades humunculares podrán irse sustituyendo por apropiados mecanismos computacionales

ya sin ningún resto irreducible de inteligencia. Véase Lycan (1981 y 1987).

R. MILLIKAN también ofrece otro tipo diferente de funcionalismo intentando llevar hasta sus últimas consecuencias la idea de que nuestros pensamientos tienen una función biológica propia y característica. Cada pensamiento en particular, cada creencia, deseo, recuerdo, percepción, etc., tendría una función biológica diferenciada. Su posición ha sido calificada de funcionalismo teleológico. Véase especialmente Millikan (1984).

Entre los autores más precavidos respecto al funcionalismo computacional, es obligado mencionar a N. BLOCK. Desde sus inicios criticó con gran agudeza muchos de sus supuestos, especialmente su indiferencia y descompromiso respecto al tipo de sistemas físicos capaces de ejemplificar las propiedades mentales. Véase en este sentido Block (1978).

Por lo que respecta a la psicología natural, S. STICH es el principal exponente de un eliminativismo de tipo computacional. Véase Stich (1983). Las interpretaciones que la psicología natural hace de nuestros procesos cognitivos serían tan arbitrarias y científicamente irrelevantes como las interpretaciones que un niño o un poeta pueden hacer de lo que esté pasando en un computador que, por ejemplo, esté manteniendo una conversación con ellos. Stich (1990) extrae de tal eliminativismo consecuencias radicales para nuestras concepciones habituales de la racionalidad.

El funcionalismo computacional clásico en todas sus versiones ha sido criticado por muy diversos autores y desde posiciones muy diferentes. H. DREYFUS (1979) argumentó que habría siempre muchas cosas que un computador, y en general una máquina construida por nosotros, no podría hacer. Dreyfus es la voz del sentido común que se rebela contra las fantasías y promesas del computacionalismo clásico y de la inteligencia artificial en sentido fuerte. También J. SEARLE se sitúa en esa posición. Su teoría de los actos de habla ocupa un lugar sumamente importante en la lingüística y la filosofía del lenguaje contemporáneas. Pero, para él, la filosofía del lenguaje es sólo una rama más de la filosofía de la mente y, paralelamente, el lenguaje una manifestación de la intencionalidad. La intencionalidad presenta características sumamente peculiares que impiden cualquier tratamiento exclusivamente formal. Así como con su teoría de los actos de habla se enfrentó al formalismo de Chomsky, en su teoría general de la intencionalidad repite la maniobra arremetiendo contra lo que considera excesos del computacionalismo y de la inteligencia artificial. Véanse

todas las referencias de Searle que señalamos en la bibliografía. Véase, también, Jacquette (1989) y LePore y van Gulick, (Eds.) (1991). Decíamos que Dreyfus y Searle se apoyan en el sentido común. Sin embargo, demasiadas veces la ciencia y la tecnología han desmentido lo que parecían ser intuiciones firmemente arraigadas en el sentido común. Y resulta estimulante darse cuenta de que algunas de esas cosas que se pensaban imposibles ya las pueden hacer nuestros computadores actuales. La distinción entre lo natural y lo artificial es, hoy día, mucho más difusa y problemática de lo que Dreyfus pensó. Y también mucho más de lo que sigue pensando Searle.

Otro frente crítico radical, esta vez contra todo intento, computacionalista o no, de abordar científicamente lo mental lo constituye T. NAGEL. Este autor representa uno de los últimos bastiones del dualismo de propiedades, incluso tal vez también de un dualismo de entidades. El punto de vista subjetivo, el punto de vista de la primera persona, nos pondría en contacto con un mundo completamente irreductible al mundo de la ciencia y del conocimiento objetivo. Los argumentos dualistas de Nagel son siempre epistemológicos. Véase Nagel (1974, 1979 y 1986). Y se asemejan mucho a los ofrecidos por F. JACKSON (1982 y 1986). Este frente crítico apela siempre a la introspección, entendida como cierta capacidad para conocer directa e inmediatamente la naturaleza misma de nuestros propios estados y procesos mentales sin recurrir a la experimentación o a otras formas de «experiencia externa». Rechazar esta concepción de la introspección no implica, sin embargo, rechazar totalmente que tenemos algún tipo de capacidad introspectiva. El problema es cómo entenderla y, si cabe, explicarla. En este sentido, son relevantes las teorías sobre la conciencia propuestas recientemente por autores como Dennett, 1991, Flanagan, 1992, Jackendoff, 1987, Lycan, 1987, y Lyons, 1986.

También S. KRIPKE (1980) ofrece fuertes argumentos dualistas, esta vez no sólo epistemológicos sino metafísicos. La intuición modal de que cualquier pretendida identidad entre las propiedades mentales y algún otro tipo de propiedades será siempre contingente es, para él, razón más que suficiente para el dualismo.

K. POPPER y J. ECCLES son asimismo dualistas. Dualistas de propiedades pero esta vez, además, claramente también de entidades. El yo no es de ninguna forma su cerebro. Lo curioso de tal posición es que pretendan que su dualismo tenga una base neurofisiológica. Casi completamente marginal al desarrollo actual de las ciencias cognitivas, merece no obstante la pena leer Popper y Eccles (1977), sobre todo sus argumentos contra el materialismo.

El CONEXIONISMO es la alternativa más global y prometedora al computacionalismo clásico. Pero es también un computacionalismo. Todo computacionalismo se desarrolla en torno a la analogía entre la mente y el computador. La mente funciona como un computador. ¿Qué clase de computador? Aquí es donde empiezan los problemas. Podemos distinguir un computacionalismo clásico, basado en modelos con una estructura computacional tipo von Neumann, y un computacionalismo de tipo conexionista, basado en modelos de procesamiento distribuido y no simbólico. Lo que está en juego es la clase de arquitectura computacional que puede tener la mente. ROSENBLATT (1959) es uno de los principales inspiradores del conexionismo. En una serie de tristes episodios dignos de la mejor sociología de la ciencia, su voz fue silenciada por Minsky y Papert (1969). Realmente, sus modelos eran bastante simples. Actualmente, sin embargo, los modelos conexionistas parecen no conocer aún el límite de sus posibilidades. Y esto, en todos los sentidos. Para una aproximación al conexionismo, veanse Bechtel & Abrahamsen (1991), Clark (1989), Horgan y Tienson, (Eds.) (1988), MCCLELLAND, RUMELHART y THE PDP RESEARCH GROUP, (1986) y SMOLENSKY (1988). Respecto a las vicisitudes de las primeras ideas conexionistas, véase Olazarán (1995).

Cercanos al conexionismo se encuentran tanto D. Dennett como Paul y Patricia Churchland. D. DENNETT es, en todos los sentidos, sumamente interdisciplinar y tremendamente imaginativo. Su afición a los experimentos mentales, a la ficción científica y filosófica, es proverbial. Véase Hofstadter y Dennett (1981). Las ideas de Dennett sobre la perspectiva intencional tienen una aplicación extensa y siempre sugerente. Al igual que sólo podemos ver las células desde la perspectiva proporcionada por un microscopio, lo mental sólo puede ser detectado y descrito desde esa perspectiva que Dennett llama intencional. Su anticartesianismo y antifodoriano le ha llevado a la elaboración de una alternativa global sobre lo mental. En esta alternativa global aparecen tratados incluso temas relativos a la libertad, a nuestro ser sujetos morales, personas, etc. El único pero gran problema con Dennett es el carácter resbaladizo e indeterminado de sus posiciones. La mayoría de las veces uno no termina de saber muy bien si la intencionalidad es o no algo real, si somos o no libres, etc. Véanse en general todas las referencias de Dennett que señalamos en el apartado bibliográfico. Véase también Dahlbom (1993).

Junto con su colega y esposa, Patricia Churchland, PAUL CHURCHLAND es actualmente uno de los más directos continuadores de las clásicas perspectivas materialistas propuestas por Armstrong (1980),

Campbell (1984), Feyerabend (1981), Smart (1962) o Rorty (1965). Y, como suele pasar, los hijos superan a los padres a la vez que heredan sus pecados. Defiende un materialismo eliminativista fuertemente ligado, por un lado, a los avances de la neurología y de los nuevos modelos computacionales conexionistas y, por otro, a una rigurosa discusión del tema de la reducción de teorías al hilo de las críticas que los planteamientos del empirismo lógico han recibido en la reciente filosofía de la ciencia. Su libro de 1979 tuvo un impacto considerable tanto en el campo de las ciencias cognitivas como en el de la filosofía de la ciencia, impacto que sus trabajos posteriores han contribuido a mantener.

PATRICIA CHURCHLAND es una crítica despiadada del computacionalismo de Fodor. Apoyándose en la neurología y en un conexionismo biológicamente orientado, su neurofilia es incluso más acusada que la de su colega y marido Paul Churchland. Nos promete continuamente una teoría unificada de la mente-cerebro. Pero, que ese sueño se convierta en realidad, resultando hasta innecesario el anterior guión, se hace depender siempre del desarrollo futuro de la ciencia. Y, como sabemos, el futuro es incierto. Todas las referencias que señalamos en el apartado bibliográfico tanto respecto a Paul como a Patricia Churchland serían de interés.

También lleno de ejemplos procedentes de la biología, la figura de F. DRETSKE se asemeja mucho a la de un explorador solitario de lo mental cuyo consejo puede resultar siempre valioso y esclarecedor. Implacable, persigue nuestra vida mental siguiendo los rastros de la información y de la causalidad. Sus trabajos sobre temas de filosofía de la mente tienen siempre una claridad y un rigor dignos del mayor respeto. Especialmente, los relativos a la constitución de los contenidos mentales a partir del flujo natural de la información y a la integración de las razones en la trama causal del mundo natural. El único problema es que muchas veces ese mundo natural parezca necesitar, a su vez, ser llenado de razones a la hora de emprender tales constituciones e integraciones. Véanse todas las referencias bibliográficas sobre Dretske que señalamos en el apartado bibliográfico. Véanse asimismo Gjelsvik (1990) y McLaughlin (Ed.) (1991).

Por el lado de las consideraciones metafísicas, no podemos acabar sin mencionar al menos a cuatro autores más. J. KIM es el filósofo actual que más se ha esforzado por precisar conceptualmente la noción de sobrevenida y explorar sus posibilidades. Asimismo destacan sus críticas al monismo anómalo de Davidson y sus análisis de la problemática noción de causación mental. En este último tema, su propuesta consiste en legitimar las relaciones causales intencionales desde la

base de relaciones causales microfísicas asociadas a sus causas y efectos mediante relaciones de sobreveniencia lo más locales y fuertes posible. La causación mental sería así sólo un tipo particular de causalidad macrofísica. Véase Kim (1993 y 1996). Otro tipo interesante de sobreveniencia sería una sobreveniencia que podríamos llamar mereológica. La intuición de que algunas de las propiedades mentales de un sujeto están determinadas por las relaciones que mantengan determinadas partes del resto de su vida mental es sugerente. E. SOSA (1987) explora justamente esta idea respecto a la propiedad de ser personas. Véase también Gomila (1995). Por otra parte, J. HAUGELAND (1982) defiende uno de los fisicalismos más débiles que se conocen, no aceptando ni siquiera una identidad de casos entre lo mental y lo físico. Lo mental estaría determinado por lo físico únicamente en el sentido de que dos mundos exactamente iguales en lo físico deben compartir todas sus propiedades mentales. Sobre las relaciones entre lo mental y lo físico sólo podríamos decir esto. Lo mental sobrevendría globalmente sobre lo físico. Pero decir algo así sería suficiente, según Haugeland, para no ver lo mental como algo separado del mundo físico. Situado en las antipodas de esto, R. RORTY defendió hace años un eliminativismo neurofisiológico radical respecto a lo mental y un total escepticismo respecto a la introspección, al pretendido carácter cualitativo de algunos estados mentales, a la incorregibilidad como marca característica de lo mental, etc. Su posición es, en líneas generales, compartida también por Armstrong (1980), Campbell (1984), Feigl (1958), Feyerabend (1963a y 1963b) o Smart (1962). Véase Rorty (1965 y 1970). De aquí, sin embargo, ha ido pasando a cierto tipo de relativismo y antirrealismo hermeneúutico donde lo mental aparece y desaparece según se adopte un lenguaje o sistema conceptual u otro. Véase, por ejemplo, Rorty (1979).

Acabaremos esta sección indicando algunas importantes antologías y compilaciones de trabajos en filosofía de la mente. Todas ellas serían de interés y gran utilidad. Son las siguientes: Blakemore & Greenfield (Eds.) (1987), Block (Ed.) (1980-1), French et al. (Comps.) (1986), Haugeland (Ed.) (1981), Horgan y Tienson (Eds.) (1988), Lycan (Comp.) (1990), Rosenthal (Comp.) (1991) y Tomberlin (Ed.) (1989 y 1990). En castellano debería tenerse en cuenta Broncano (Ed.) (1995b) y Rabossi (Comp.) (1995).

## 5. Algunos problemas

¿Cuáles son los problemas centrales de la actual filosofía de la mente? No vamos a responder a esta pregunta. Centralidad y periferia

tienen una localización muy variable en un área de conocimiento en continuo desarrollo. Sin embargo, sí mencionaremos y comentaremos varios grupos importantes de problemas abiertos.

Existe, en primer lugar, un amplio grupo de problemas que podemos llamar *estructurales*. Y tienen que ver tanto con el nivel general de análisis que debería considerarse apropiado para el tratamiento de lo mental como con la particular arquitectura cognitiva que habría que tener presente. En cuanto al *nivel de análisis*, las opciones varían desde considerar que el nivel apropiado ha de ser el neurofisiológico hasta considerar que el único nivel desde el cual podemos realmente identificar fenómenos mentales es un nivel interpretativo irreducible. Entre medias, se situarían los niveles de análisis estímulo-respuesta propios del conductismo y los niveles funcionales característicos del computacionalismo clásico y del conexionismo. Y, también, la mayoría de los niveles anti-individualistas sugeridos por aquellos que defienden la necesidad de tener en cuenta los entornos de los sujetos o su historia evolutiva y cultural. La *arquitectura cognitiva* puede ser incluso mucho más variada. Puramente neurofisiológica, de «caja negra», computacional clásica o conexionista, etc., existiendo dentro de cada una de estas arquitecturas cognitivas generales decisivas diferencias de detalle.

En sí mismos, los problemas del nivel de análisis y de la arquitectura cognitiva son dos tipos de problemas distintos. La elección del nivel de análisis es fundamentalmente metodológica. La postulación de una arquitectura cognitiva concreta supone ya, en cambio, la elaboración de hipótesis sustantivas. Pero las respuestas que se ofrezcan al primer tipo de problemas restringen enormemente las respuestas que cabe dar a los problemas del segundo tipo. Un partidario de niveles de análisis neurofisiológicos, por ejemplo, no puede aceptar arquitecturas de «caja negra». Y, por supuesto, un partidario de niveles de análisis irreduciblemente interpretativos no puede aceptar ninguna arquitectura cognitiva. Hay en todo esto, sin embargo, una posibilidad muy atrayente que permanece abierta. Se trata de la posibilidad de que lo mental requiera no un único y exclusivo nivel de análisis ni una única y exclusiva arquitectura cognitiva, sino tratamientos siempre *híbridos*. Ciertas parcelas de lo mental podrían requerir así, por ejemplo, niveles de análisis neurofisiológicos, y arquitecturas cognitivas acordes con esa elección metodológica, otras parcelas de lo mental podrían requerir análisis preferentemente conductistas, o funcionales, etc. Distinguir esto del simple eclecticismo no es fácil. Pero, sin embargo, puede que sí sea lo más razonable.

Otro grupo relevante de problemas surge en torno a las *propiedades semánticas* que habitualmente atribuimos a gran parte de nuestros pensamientos. ¿En qué consisten, de dónde proceden y cómo se constituyen propiedades semánticas como el significado, la referencia y la verdad? Esta pregunta es particularmente pertinente si uno admite que el origen de todas las propiedades semánticas que encontramos en nuestros lenguajes naturales, en cualquier acto comunicativo en general, debe situarse en último término en la actividad mental de ciertos sujetos al interactuar con sus entornos. La percepción y el aprendizaje de un lenguaje parece que deberían jugar aquí un papel importante. Aunque no tratadas al estilo del conductismo o de la neurofisiología. Algún tipo de funcionalismo computacional parece aquí lo más razonable. Algún tipo de computacionalismo que explique las propiedades semánticas de nuestros pensamientos en base a nociones informacionales. Sin embargo, desde un punto de vista computacional puramente individualista, ya sabemos que, por decirlo en términos metafóricos, no habría diferencia alguna entre ser un robot y ser un computador que sueña ser un robot. Y esto sería estrictamente aplicable tanto al computacionalismo clásico como al conexionista. Un sistema cognitivo computacional, ya sea clásico o conexionista, puede estar tan separado del mundo externo como una conciencia cartesiana. Acaso un punto de vista computacional pero no-individualista pudiera arreglar las cosas. Sin embargo, no tenemos ninguna idea clara y precisa de lo que sería tal punto de vista.

Y si detrás de los problemas semánticos hemos visto reaparecer el clásico problema del mundo externo (del «anclaje» que puede tener nuestra mente en un mundo que no sólo es mental, que no sólo es un mundo representado), también hay otro grupo de problemas en la actual filosofía de la mente que recuerdan otro problema igualmente muy clásico: el de las *otras mentes*. Y esto, en dos sentidos. De una parte, se encuentran los conocidos problemas relativos a separar, dentro de todo aquello que puede llegar a ejemplificar una determinada estructura funcional, aquellas genuinas ejemplificaciones cognitivas de otras cosas que nos parecen carentes de cognición. Esto también afecta tanto al funcionalismo computacionalista clásico como al funcionalismo conexionista y a cualquier otra variedad no computacional de funcionalismo. Aunque las propiedades mentales sean en gran medida múltiplemente realizables respecto a sus sustratos físicos, parece que debe haber algo más. Algo que ponga límites a esa múltiple realizabilidad. Pero no sabemos muy bien qué puede ser ese algo más.

De otra parte, están los problemas relacionados con la atribución de propiedades mentales a sujetos no humanos (animales, máquinas,

marcianos, etc.) y a sujetos humanos con capacidades mentales no normales o no completamente desarrolladas (personas con lesiones cerebrales, con incapacidades o trastornos psicológicos, niños preverbales, etc.). El atribuir o no ciertas propiedades mentales (por ejemplo, la capacidad de sentir realmente dolor) tiene consecuencias prácticas importantes. Y esto nos obliga a ser muy cautos. Si hace un momento necesitábamos poner Límites la múltiple realizabilidad de lo mental, ahora parece que lo más razonable es no imponer condiciones demasiado restrictivas. Aunque, nuevamente, seguimos sin saber muy bien en qué punto quedarnos.

Lo que acabamos de señalar nos conduce directamente a otro tipo de problemas. Nuestra mente, la mente humana, ha querido siempre distinguirse por ser capaz de tener *conciencia* (y auto-conciencia). Pero la conciencia ha sido para la filosofía contemporánea de la mente una especie de oveja negra. Una oveja negra, además, sospechosa de ser en realidad un lobo disfrazado. La sensación de que eran demasiados los problemas que se escondían detrás del concepto de conciencia, y el convencimiento de que la mayoría de ellos estaban mal planteados, hacía preferible dirigir las investigaciones por otros derroteros. Implícita o explícitamente se pensaba que cuando ya se dispusiera de la suficiente claridad conceptual y de un buen conocimiento de los detalles empíricos en las otras áreas de la mente, sólo entonces podría ser abordable con seriedad el tema de la conciencia.

Inocente oveja negra o peligroso lobo disfrazado, el caso es que le está llegando por fin el turno a la conciencia. Comenzamos a tener teorías sobre la conciencia. Esto es, sin duda, esperanzador. Pero, no obstante, con la conciencia una y otra vez acaba repitiéndose el mismo problema. Se llene como se llene de contenido explicativo un enunciado del tipo «ser consciente consiste en ...», siempre nos queda una especie de insatisfacción. Sabemos que esto no es razonable. Y que no ocurre con ninguna otra clase de explicaciones. Quedarse siempre insatisfecho se llene como se llene de contenido explicativo una teoría de la conciencia es algo que no tiene justificación. Y sin embargo sucede.

Junto con la conciencia, la *racionalidad* es otro de nuestros bienes más preciados. Sin suponer algún tipo de racionalidad no hay forma siquiera de identificar sujetos psicológicos. El mismo hecho de disponer de un sistema cognitivo parece presuponer que somos racionales. Ahora bien, ¿cuánto de racionales? Aquí es donde empiezan los problemas. Somos sujetos finitos. Exigir una racionalidad ideal es pedir un imposible. Nuestras capacidades inferenciales son bastante pobres, la ordenación de nuestras preferencias es todo lo que se quiera menos

coherente, nuestra memoria (al menos nuestra memoria a largo plazo) es un gran almacén de inconsistencias, etc. Lo más razonable parece ser exigir sólo una racionalidad mínima. Pero, ¿dónde está la diferencia entre tal racionalidad mínima y la falta de racionalidad?

Nuestros pensamientos no sólo son razones de nuestra acción. También parecen ser su causa. El problema de la *causación mental* consiste justamente en el problema de entender cómo las razones pueden ser también causas en un mundo donde sólo quisieramos admitir causas físicas (o causas muy directamente vinculadas a lo físico). Ciertamente, nos atrae la idea de considerar que nuestros estados mentales son a la vez razón y causa de cosas como nuestras acciones. Pero ambas características parecen excluirse mutuamente. Son muchas las discusiones originadas en torno a este tema. Y son muchas las propuestas sugeridas. Pero en todas ellas nos vemos abocados siempre al mismo dilema: eliminativismo o dualismo. O las propiedades mentales acaban diluyéndose en lo físico o acaban constituyendo un mundo aparte. Parece como si la armonía entre razones y causas fuera tan necesaria como imposible de conseguir.

Son, como se ve, muchos los problemas pendientes. Otra manera de verlo es decir que la filosofía de la mente realmente está muy viva. Y que nos está enseñando una nueva forma de hacer filosofía. Se trata de una filosofía que pone constantemente en práctica la no existencia de fronteras claramente definidas entre filosofía y ciencia. Pero, a la vez, es una filosofía que no aspira a proporcionar más conocimientos, o conocimientos radicalmente diferentes, de los conocimientos conseguidos a través de la ciencia. A lo que aspira la filosofía de la mente es a comprender mejor lo que vamos conociendo sobre la mente. Sin embargo, es justamente en este punto donde surge un último problema inquietante que es necesario señalar. Y una terrible sospecha. Al examinar todos los anteriores problemas acabábamos siempre considerando como razonable algo que, no obstante, quedaba aún muy lejos de nuestro alcance. Y aquí está el problema inquietante. Y la terrible sospecha de que, después de todo, tal vez seamos *más complejos que inteligentes*.

### Bibliografía

- ARMSTRONG, D. (1980): *The Nature of Mind*. Univ. of Queensland Press, Queensland.  
ECHTEL, W. y A. ABRAHAMSEN (1991): *Connectionism and the Mind. An Introduction to Parallel processing in networks*. Cambridge, Basil Blackwell.

- BLAKEMORE C. y GREENFIELD, S. (Eds.) (1987): *Mindwaves: thoughts on intelligence, identity and consciousness*. Basil Blackwell, New York.
- BLOCK, N. (1978): «Troubles with Functionalism», en SAVAGE (Ed.) (1978): [También en ROSENTHAL (Ed.) (1991)].
- (Ed.) (1980-1): *Readings in the philosophy of Psychology*. 2 vols. Cambridge Univ. Press, Cambridge.
- BRONCANO, F. (Ed.) (1995a): «Nuevas meditaciones sobre la técnica. Trotta, Madrid.
- (Ed.) (1995b): *La mente humana*, Enciclopedia Ibero-Americana de Filosofía, vol. 9. Trotta, Madrid.
- BURGE, T. (1979): «Individualism and the Mental», en FRENCH et al. (Eds.) (1979).
- (1986): «Individualism and psychology» *Philosophical Review*, 95.
- BUTLER, R. (Ed.) (1965): *Analytical Philosophy, Part II*. Basil Clackwell, London.
- CAMPBELL, K (1984): *Body and Mind*. Doubleday, New York.
- CARNAP, R. (1937): «Psychology in physical language». *Erkenntnis*, 3.
- CLARK, A. (1989): *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. MIT Press, Cambridge.
- CHERNIAK, C. (1986): *Minimal Rationality*. MIT Press, Cambridge.
- CHOMSKY, N. (1957): *Syntactic Structures*. The Hague, Mouton [Trad. cast. Estructuras Sintácticas. S. XXI, Madrid, 1974].
- (1959): «Review of Skinner's "Verbal Behavior"», *Language*, 35,
- (1965): *Aspects of the Theory of Syntax*. MIT Press. Cambridge. [Trad. cast. *Aspectos de la teoría de la sintaxis*]. Aguilar, Madrid.
- (1980): *Rules and Representations*. Colubia Univ. Press. New York. [Trad. cast. *Reglas y representaciones*. FCE, México, 1983].
- (1985): *Knowledge of Language: Its Nature. Origins and Use*. Praeger, New Yor. [Trad. cast. *El conocimiento del lenguaje. Su naturaleza, origen y uso*: Alianza, Madrid, 1989].
- CHURCHLAND, Patricia (1978): «Fodor on Language Learning». *Synthese*, 38.
- (1980): «Language, Thought, and Information Processing», *Nous*, 14.
- (1982): «Mind-Brain reduction: New light from the philosophy of science», *Neuroscience*, 7/5.
- (1986): *Neurophilosophy. Toward a Unified Science of the Mind/Brain*. MIT Press, Cambridge.
- CHURCHLAND, Paul (1979): *Scientific realism and the plasticity of mind*. Cambridge, Univ. Press, Cambridge.
- (1981): «Eliminative Materialism and the Propositional Attitudes», *The Journal of Philosophy* LXXVIII, 2.
- (1983): «Folk Psychology and the Explanation of Human Behavior», *Philosophical Perspectives*, 3.
- (1984): *Matter and Consciousness: A contemporary introduction to the philosophy of mind*. MIT Press, Cambridge.
- (1985): «Reduction, qualia, and the direct introspection of brain states», *Journal of Philosophy*, 82.
- (1989): *A Neurocomputational Perspective. The Nature of Mind and the Structure of Science*. MIT Press, Cambridge.
- CHURCHLAND, Patricia y CHURCHLAND, Paul (1983): «Stalking the Wild Epistemic Engine». *Nous*, 17.

- CHURCHLAND, Paul y CHURCHLAND, Patricia (1983): «Functionalism, qualia, and intentionality». *Philosophical Topics*, 12.
- (1990): «¿Podría pensar una máquina?». *Investigación y ciencia*. Marzo.
- DAHLBOM, B. (Ed.) (1993): *Dennett and his Critics*. Basil Blackwell, Cambridge.
- DAVIDSON, D. (1963): «Actions, Reasons and Causes», *Journal of Philosophy*, 60. [También en DAVIDSON (1980)].
- (1970): «Mental Events», en FOSTER y SWANSON (Eds.) (1970) [También en DAVIDSON (1980)].
- (1980): *Essays on Actions & Events*. Oxford Univ. Press, New York.
- (1984): *Inquiries into Truth and Interpretation*. Oxford Univ. Press, New York.
- DENNETT, D. (1969): *Content and Consciousness*. Routledge and Kegan Paul, London.
- (1971): «Intentional Systems». *Journal of Philosophy*, 68, n. 4 [También en DENNETT (1978)].
- (1978): *Brainstorm: Philosophical essays on mind and Psychology*. Bradford Book, Montgomery. [También en MIT Press, Cambridge].
- (1984): *Elbow room: The varieties of free will worth wanting*. MIT Press, Cambridge.
- (1986): «The Logical Geography of Computational Approaches: a View from the East Pole», en HARNISH y BRAND (Eds.) (1986).
- (1987): *The Intentional Stance*. MIT Press, Cambridge.
- (1990a): «The Interpretation of Texts, People and Other Artifacts», *Philosophy and Phenomenological Research*, vol. L, Supplement, Fall.
- (1990b): «Real Patterns». *Journal of Philosophy*, 89.
- (1990c): «Memes and the Exploitation of Imagination». *The Journal of Aesthetics and Art Criticism*, 48:2.
- (1991): *Consciousness Explained*. Little, Brown and Company. Boston.
- (1996): *Kinds of Minds*. Basic Books, New York.
- DRETSKE, F. (1980): «The Intentionality of Cognitive States». *Midwest Studies in Philosophy*, V. [También en ROSENTHAL (1991)].
- (1981): *Knowledge and the Flow of Information*. MIT Press, Cambridge.
- (1983): «Reasons and Causes». *Philosophical Perspectives*, 3.
- (1986): «Misrepresentation», en BOGDAN (Ed.) (1986).
- (1988): *Explaining Behavior. Reasons in a World of Causes*. MIT Press, Cambridge.
- DREYFUS, H. (1979): *What Computers Can't Do. A Critique of Artificial Reason*. Harper & Row, New York.
- FEIGL, H. (1958): *The «Mental» and the «Physical». The Essay and a Postscript*. Univ. of Minnesota Press. Minneapolis.
- FEYERABEND, P. (1963a): «Materialism and the Mind-Body Problem». *Review of Metaphysics*, XVII.1, 65.
- (1963b): «Mental events and the brain». *Journal of Philosophy*, 60.
- FIELD, H. (1978): «Mental representation». *Erkenntnis*, 13 [También en BLOCK (Ed.) (1981)].
- FLANAGAN, O. (1992): *Consciousness Reconsidered*. MIT Press, Cambridge.
- FODOR, J. (1965): *Psychological Explanation*. Random House. Nueva York. [Trad. Cast.: La explicación psicológica. Cátedra, Madrid, 1984].
- (1974): «Special Sciences», *Synthese*, 28 [recogido también en FODOR (1981)].
- (1975): *The Language of Thought*. Harper & Row, Nueva York. [Trad. Cast.: El lenguaje del Pensamiento. Alianza, Madrid, 1984].
- (1978): «Propositional Attitudes». *The Monist*, 59 [También en FODOR (1981)].

- (1980): «Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology». *The Behavioral and Brain Sciences* III, 1.
- (1981): *Representations*. MIT Press, Cambridge.
- (1983): *The Modularity of Mind*. MIT Press, Cambridge. [Trad. Cast.: *La modularidad de la mente*. Morata, Madrid, 1986].
- (1985): «Fodor's guide to mental representation: The intelligent auntie's vade-mecum». *Mind*, 94 [También en FODOR (1990b)].
- (1987a): *Psychosemantics. The Problem of Meaning in the Philosophy of Mind*. MIT Press, Cambridge.
- (1987b): «Why There Still has to be a Language of Thought», apéndice de FODOR (1987a).
- (1990a): «Psychosemantics or: Where Do Truth Conditions Come From?», en LYCAN (Ed.) (1990).
- (1990b): *A Theory of Content and Other Essays*. MIT Press, Cambridge.
- (1994): *The Elm and the Expert*. MIT Press, Cambridge.
- FODOR, J. y PYLYSHYN, Z. (1988): «Connectionism and the Mental». *Cognition*, 28.
- FRENCH, P. et al. (Eds.) (1979): *Midwest Studies in Philosophy*, vol. 4. Univ. of Minnesota Press, Minneapolis.
- FRENCH, P. et al. (Eds.) (1979): *Midwest Studies in Philosophy*, vol. 10: *studies in the Philosophy of Mind*. Univ. of Minnesota Press, Minneapolis.
- GARCÍA-ALBEA, J. (1986): *Percepción y computación*. Pirámide, Madrid.
- GJELSVIK, O. (1990): «Representational Content and the Esplanationof Behavior», *Inquiry*, 33.
- GOMILA, T. (1995): «From Cognitive Systems to Persons», en GLYMOUR, C.; HAYES, P. y FORD, K. (Eds.) (1995).
- GLYMOUR, C.; HAYES, P. y FORD, K. (Eds.) (1995): *Android Epistemology*. MIT Press, Cambridge.
- GRICE, H. (1957): «Meaning». *Philosophical Review*, 66.
- (1969): «Utterer's Meaning and Intentions», *Philosophical Review*, 78.
- GUNDERSON, K. (Ed.) (1975): *Minnesota studies in the Philosophy of science*, 7. Univ. of Minnesota Press, Minneapolis.
- HARMAN, G. (1973): *Thought*. Princenton Univ. Press, Princenton.
- HAUGELAND, J. (Ed.) (1981): *Mind Design*. MIT Press, Cambridge.
- (1982): «Weak Supervenience». *American Philosophical Quarterly*, 19.
- (1985): *Artificial Intelligence: The Very Idea*. MIT Press, Cambridge.
- HOFSTADTER, D. y DENNET, D. (1981): *The Mind's I: Fantasies and Refiections on Mind and Soul*. Basic Books, New York.
- HOLLAND, J. et al. (1989): *Induction. Processes of Inference. Learning and Discovery*. MIT Press, Cambridge.
- HORGAN, T. y TIENSON, J. (Eds.) (1988): *Spindel Confeence 1987: Connectionism and the Philosophy of Mind*. *Southern Journal of Philosophy*, XXVI, supplement.
- JACKENDOFF, R. (1987): *Consciousness and the Computational Mind*. MIT Press, Cambridge.
- JACKSON, F. (1982): «Epiphenomenal qualia». *Philosophical Quarterly*, 32.
- (1986): «What Mary Didn't Know». *The Journal of Philosophy*, 83.
- JACQUETTE, D. (1989): «Searle's Intentionality Thesis». *Synthese*, 80.
- JOHNSON-LAIRD, P. (1983): *Mental Models: Towards a Cognitive Science of Language, Inference and Consiousness*. Cambridge Univ. Press, Cambridge.

- KIM, J. (1993): *Supervenience and Mind*. Cambridge Univ. Press, Cambridge.
- (1996): *Philosophy of Mind*. Westview Press, Colorado.
- KRIPKE, S. (1980): *Naming and Necessity*. Cambridge Univ. Press, Cambridge.
- LEPORE, E, (Ed.) (1986): *Truth and Interpretation. Perspectives on the Philosophy of Donald Davidson*. Basil Blackwell, New York.
- LEPORE, E, y MCLAUGHLIN, B. (Eds.) (1985): *Actions & Events: Perspectives on the Philosophy of Donald Davidson*. Basil Blackwell, New York.
- LEPORE, E, y VAN GULICK, R. (Eds.) (1991): *John Searle and his Critics*. Basil Blackwell, Cambridge.
- LEWIS, D. (1972): «Psychophysical and Theoretical Identifications», *Australasian Journal of Philosophy*, 50.
- (1980): «Mad Pain and Martian Pain», en BLOCK (Ed.) (1980) [También en ROSENTHAL (1991)].
- LOEWER, B. y REY, G. (Eds.) (1991): *Meaning in Mind. Fodor and his Critics*. Basil Blackwell, Cambridge.
- LYCAN, W. (1981): «Toward a homuncular theory of believing». *Cognition and Brain Theory*, 4.
- (1987): *Consciousness*. MIT Press, Cambridge.
- (Comp.) (1990): *Mind and Cognition*. Basil Blackwell, Cambridge.
- LYONS, W. (1986): *The Disappearance of Introspection*. MIT Press, Cambridge.
- MCCLELLAND, J., RUMELHART, D., y The PDP Research Group (1986): *Parallel distributed Processing: Explorations in the microstructure of cognition*. 2 vols. MIT Press, Cambridge.
- MCGINN, C. (1989): *Mental Content*. Basil Blackwell, New York.
- MCLAUGHLIN, B. (Ed.) (1991): *Dretske and his Critics*. Basil Blackwell, Cambridge.
- MARR, D. (1982): *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman, New York. [Trad. cast. La visión. Alianza, Madrid, 1985].
- MILLIKAN, R. (1984): *Language, Thought, and Other Biological Categories*. MIT Press, Cambridge.
- MINSKY, M. y PAPERT, S. (1969): *Perceptrons*. MIT Press, Cambridge.
- NAGEL, T. (1974): «What is it like to be a bat?». *Philosophical Review*, 83.
- (1979): *Mortal Questions*. Cambridge Univ. Press, Cambridge.
- (1986): *The View from Nowhere*. Oxford Univ. Press, Oxford.
- NISBETT, R. y ROSS, L. (1980): *Human inference: Strategies and Shortcomings of social judgment*. Prentice-Hall, New Jersey.
- OLAZARAN, M. (1995): «Controversias y estructuración social de las comunidades científico-tecnológicas: un estudio de caso en inteligencia artificial», en BRONCANO (Ed.) (1995a).
- PETTIT, Ph. y MCDOWELL, J. (Eds.) (1986): *Subject, Thought and Context*. Clarendon Press, Oxford.
- POPPER, K. y ECCLES, J. (1977): *The self and its brain*. Berlin, Springer [Trad. Cast. *El yo y su cerebro*. Labor, Barcelona, 1980].
- PUTNAM, H. (1960): «Minds and machines», en HOOK (Ed.) (1960).
- (1965): «Brains and Behavior», en BUTLER (Ed.) (1965).
- (1967): «The Nature of Mental States», en LYCAN (Ed.) (1990).
- (1975a): «The meaning of meaning», en GUNDERSON (Ed.) (1975). [También en PUTNAM (1975b)].

- (1975b): *Mind. Language and Reality. Philosophical Papers II*. Cambridge Univ. Press, Cambridge.
- (1988): *Representation and Reality*. MIT Press, Cambridge.
- (1992): *Renewing Philosophy*. Harvard Univ. Press, Cambridge.
- PYLYSHYN, Z. (1984): *Computation and Cognition: toward a foundation for cognitive science*. MIT Press, Cambridge [Trad. Cast. Computación y conocimiento. Hacia una nueva fundamentación de la ciencia cognitiva. Debate, Madrid, 1988].
- RABOSI, E. (Comp.) (1995): *Filosofía de la mente y ciencia cognitiva*. Paidós, Barcelona.
- RIVIERE, A. (1991): *Objetos con mente*. Alianza, Madrid.
- RORTY, R. (1965): «Mind-Body Identity, Privacy, and Categories», *Review of Metaphysics*, XIX.1, 73.
- (1970): «Incorrigibility as the Mark of the Mental», *Journal of Philosophy*, 67.
- (1979): *Philosophy and the Mirror of Nature*, Princeton, Univ. Press, Princeton [Trad. cast. *La Filosofía y el espejo de la naturaleza*. Cátedra, Madrid, 1983].
- ROSLNBLATT, F. (1959): *Principles of Neurodynamics*. Spartan Books, New York.
- ROSENTHAL, D. (Comp.) (1991): *The Nature of Mind*. Oxford Univ. Press, Oxford.
- RYLE, G. (1949): *The Concept of Mind*. Hutchinson, London.
- SAVAGE, C. (Ed.) (1978): *Minnesota Studies in the Philosophy of Science vol. 9*. Univ. of Minnesota Press, Minneapolis.
- SCHIFER, S. (1987): *Remnants of Meaning*. MIT Press, Cambridge.
- SEARLE, J. (1980): *Minds, Brains, and Programs*, *Behavioral and Brain Sciences*, 3.
- (1983): *Intentionality: An Essay in the Philosophy of Mind*. Cambridge Univ. Press, Cambridge.
- (1984): «Intencionalidad y su lugar en la naturaleza». *Synthese*, 61.
- (1985): *Minds, Brains and Science*. Harvard Univ. Press, Cambridge [*Mentes, cerebros y ciencia*. Cátedra, Madrid, 1985].
- (1987): «Mind and Brains without Programs», en BLAKEMORE y GREENFIELD (Eds.) (1987).
- (1990): «¿Es la mente un programa informático?». *Investigación y ciencia*, marzo.
- (1992): *The Rediscovery of the Mind*. MIT Press, Cambridge.
- SELLARS, W. (1956): «Empirism and the Philosophy of Mind» en SELLARS (1963).
- (1963): *Science, perception and reality*. Routledge and Kegan Paul, London.
- SHANNON, C. (1949): *The Mathematical Theory of Communications*. Univ. of Illinois Press, Champaign-Urbana [Introducción de Warren Weaver].
- SKINNER, B. (1933): *Science and Human Behavior*, Mcmillan.
- (1957): *Verbal Behavior*. Appleton, New York. [Trad. cast. *Conducta verbal*. Trillas, México, 1981].
- SMART, J. (1962): «Sensations and Brain Processes», *Philosophical Review*, 68.
- SMOLENSKY, P. (1988): «On the proper treatment of connectionism» *Behavioral and Brain Sciences*, 11.
- SOSA, E. (1987): «Subjects Among Other Things», en TOMBERLIN (Ed.) (1987).
- STICH, S. (1983): *From Folk Psychology to Cognitive Science: The case against belief*. MIT Press, Cambridge.
- (1990): *The Fragmentation of Reason. Preface to a Pragmatic Theory of Cognitive Evaluation*, MIT Press, Cambridge.
- TOMBERLIN, J. (Ed.) (1987): *Philosophical Perspectives. 1: Metaphysics*. Ridgeview Publishing Company, Atascadero (California).
- (Ed.) (1989): *Philosophical Perspectives. 3: Philosophy of Mind and Action Theory*. Ridgeview Publishing, Atascadero (California).

- (Ed ) (1987): *Philosophical Perspectives. 3: Philosophy of Mind and Action Theory*. Ridgeview Publishing, Atascadero (California).
- TURING, A. (1950): «Computing machinery and intelligence», *Mind*, 59.
- VEGA, M. (1984): *Introducción a la Psicología Cognitiva*. Alianza, Madrid.
- UNGER, P. (1990): *Identity. Consciousness and Value*. Oxford Univ. Press. New York.
- WATSON, J. (1913): «Psychology as the behaviourist views it». *Psychological Review*, 20.
- (1930): *Behaviorism*, New York, Norton [Trad. cast. *El conductismo*. Paidós, Buenos Aires, 1972].
- WOODFIELD A. (1982): *Thought and Object. Essays on Intentionality*. Oxford Univ. Press, New York.