

Filosofía de la Mente: el estado de la cuestión

(Philosophy of Mind: the State of the Art)

Josep L. PRADES

BIBLID [0495-4548 (2006) 21: 57; pp. 315-332]

Uno de los síntomas de la falta de cohesión en una tradición cultural es la dispersión: la ruptura de los supuestos compartidos que definen tanto la forma aceptable de los problemas como la relevancia de las posibles respuestas. Si la filosofía contemporánea es un ejemplo de esa situación, la filosofía de la mente es una de las áreas donde la fragmentación cultural se muestra de una forma más contundente. En ciertas zonas de la discusión filosófica (como la filosofía del lenguaje, la metafísica o la epistemología) existen desacuerdos fundamentales, pero, tales desacuerdos aparecen todavía explícitamente en la controversia contemporánea. Por citar un caso: la discusión entre las concepciones neo-fregeanas y las radicalmente anti-fregeanas es una discrepancia fundamental en filosofía del lenguaje. Sin embargo, no nos resulta extraño que la discrepancia en cuestión sea uno de los objetos explícitos de reflexión en la filosofía del lenguaje actual.

No sucede lo mismo en filosofía de la mente. Aquí, las diferencias fundamentales son el resultado de divergencias más básicas en otras áreas, como filosofía del lenguaje o metafísica. De esas discrepancias surgen percepciones muy diferentes sobre la forma misma de los problemas a tratar. Como consecuencia, en muchos casos, parece que la definición misma del problema es ininteligible para las concepciones alternativas. Por ejemplo, la idea misma de cuáles deben ser las cuestiones fundamentales en filosofía de la percepción depende de consideraciones muy generales sobre el contenido intencional, que, a su vez, están vinculadas a una toma de posición en cuestiones básicas de filosofía del lenguaje y de metafísica. La reacción típica del defensor de una concepción computacional de la percepción ante las propuestas del realismo directo neo-gibsoniano es la de que su adversario se niega a responder al problema básico: ¿cómo es posible que el cerebro reconstruya un rico contenido perceptivo que va más allá de los estímulos causalmente relevantes que impactan sobre el sistema nervioso? La reacción del adversario es la de insistir en que el enunciado mismo del supuesto problema es ininteligible, que depende de confusiones conceptuales. No es que no haya un territorio en que se discuta sobre esta discrepancia fundamental. Lo hay, pero pertenece a áreas de la discusión filosófica que tradicionalmente no son reconocidas como patrimonio de la filosofía de la mente.

Todo esto afecta al formato que debe tener el presente comentario. Por una parte, la panorámica presentada no puede aislarse de controversias y discusiones fundamentales que pertenecen a otras áreas más básicas de la reflexión filosófica e, incluso, a la concepción misma de la filosofía y de sus relaciones con la ciencia. Por otra parte, los



límites obvios de espacio impiden tratar de dibujar con precisión tales relaciones de dependencia. La estrategia escogida ha sido la de esbozar, en primer lugar, la que considero que es la oposición fundamental que recorre la literatura filosófica actual sobre la mente. Y, en segundo lugar, me preocuparé de apuntar ciertas conexiones entre tal divergencia básica y el tratamiento diferencial de ciertos problemas particulares: causalidad mental, percepción, intencionalidad, conciencia. Naturalmente, tendré que presuponer cierta familiaridad previa del lector con las taxonomías más generales que han rotulado las discusiones en filosofía de la mente en las últimas décadas (funcionalismo analítico, psico-funcionalismo, teorías de la identidad de tipos, teorías de la identidad de casos) para tratar de reconstruir las líneas maestras de una controversia que, muchas veces, cruza las fronteras que parecerían presuponerse en tales taxonomías.

1. *La herencia del conductismo lógico*

La etiqueta de “conductismo lógico” ha desaparecido prácticamente como el nombre de una posición respetable. El *conductismo lógico* pretendía la reducción de los estados mentales a disposiciones conductuales: un deseo o una creencia eran analizados como meras disposiciones a actuar de cierta forma cuando se dan ciertas condiciones. Una crítica habitual al conductismo lógico ha sido la de que olvidaba tanto el rol causal de los estados internos del organismo como el hecho de que muchas actitudes proposicionales (por ejemplo, creencias sofisticadas) no poseen manifestaciones conductuales características. El *funcionalismo* puede considerarse como un heredero natural del conductismo, con la salvedad de que trata de evitar estos dos problemas básicos. Un tipo de estado mental, para el funcionalismo, se define en términos de cierto rol funcional. Pero el rol funcional no es necesariamente una disposición a cierta forma de conducta manifiesta y puede definirse a través de multitud de relaciones con otros estados internos del organismo. Sea como fuere, es importante observar una discrepancia fundamental en el uso actual de la etiqueta “conductismo”: una discrepancia que explica el hecho de que las posiciones que todavía son acusadas de ser conductistas puedan permitirse negar la acusación.

El caso arquetípico serían las *concepciones neo-wittgensteinianas de la mente*. Estas concepciones suponen que un organismo cuya conducta manifiesta y observable fuera indistinguible de la conducta de un ser humano, de un ser al que consideramos como paradigmáticamente ejemplificando estados mentales, sería un organismo que también ejemplificaría estados mentales. Este compromiso es relativamente independiente de cuál sea la organización interna del organismo en cuestión. Seres con estructura material muy diferente tendrían que contar como dotados de mentalidad en la medida en que satisficieran los criterios ordinarios que rigen nuestras atribuciones en tercera persona, tal y como esos criterios son satisfechos paradigmáticamente por los miembros de la especie humana. ¿En qué sentido esa concepción general de la mente cuenta o no como conductismo lógico? Sus adversarios suelen insistir en que lo es. Sus defensores, lo niegan. La diferencia fundamental de interpretación estriba en el peso que se dé a la cuestión de la *reducción*: esta concepción de la mente no tiene por qué estar comprometida con una reducción de los predicados mentales a disposiciones conduc-

tuales o roles funcionales, caracterizados en términos no intencionales. Ésta es la diferencia fundamental tanto respecto al programa tradicional del conductismo lógico como a las versiones más habituales de funcionalismo. Por supuesto, esa diferencia puede ser acusada de inconsistente, o de tratar de ocupar una posición constitutivamente inestable. La discrepancia se sitúa necesariamente en un territorio mucho más general: ¿es inteligible la existencia de ciertos vínculos a priori —entre ciertas formas de comportamiento y la presencia de estados mentales— sin que los tipos mentales sean reducibles a tipos no intencionales? Hay, pues, una concepción general de la mente humana que puede ser descrita, en líneas generales, como guardando ciertas similitudes con el conductismo lógico y que, en modo alguno, ha desaparecido de la discusión contemporánea. Puede rastrearse en la producción filosófica de los seguidores de Wittgenstein, o de filósofos como Daniel Dennett, Donald Davidson, o el último Hilary Putnam. Para ellos, los criterios ordinarios de atribución de mentalidad están constitutivamente vinculados a la propiedad atribuida, de tal modo que casos paradigmáticos de satisfacción de tales criterios no dejan abierta cuestión alguna sobre si el sujeto atribuido tiene o no estados mentales. Ello, por supuesto, no implica la tesis de que la denominada “psicología popular” es una teoría científica, ni la tesis de que sus generalizaciones son asimilables a leyes que individualizan los poderes causales de las propiedades mentales ni, por supuesto, la tesis de que sea posible reducir los significados de las atribuciones de mentalidad a un vocabulario no psicológico.

No es extraño, sin embargo, que tal concepción general de la mente sea asimilada a una variante de funcionalismo que hoy en día ha pasado a ser minoritaria: el denominado “funcionalismo analítico”. El *funcionalismo analítico* se caracteriza normalmente como una tesis sobre el significado de los términos mentales. Considera que nuestros predicados mentales ordinarios pueden ser analizados en términos de relaciones causales entre estímulos, respuestas y otros estados mentales. La restricción que le impone el adjetivo “analítico” es la de que tal reducción está implícita en nuestra competencia ordinaria en el uso de los predicados mentales. Es posible, por supuesto, descubrir empíricamente leyes complejas sobre nuestra psicología a las que sólo nos da acceso el conocimiento científico sofisticado. Un funcionalista analítico tiene que insistir, sin embargo, en que tal tipo de leyes pueden ser violadas por psicologías alternativas sin que eso cuente en contra de la ejemplificación de los estados mentales relevantes. Las leyes que no pueden ser violadas, en ningún mundo posible, son aquellas que forman parte de los criterios de competencia lingüística que rigen las atribuciones ordinarias de mentalidad.

Es importante, para entender muchas de las discusiones contemporáneas, tener presente qué diferencia el funcionalismo analítico de las concepciones de la mente que caracterizaré como no-reduccionistas. En esta presentación, considero que las formas habituales de funcionalismo, incluso las formas de funcionalismo a priori, cuentan como *reduccionistas* en la medida en que se comprometen con la reducción del vocabulario intencional a tipologías funcionales que, en si mismas, son inteligibles sin presuponer nociones intencionales. Se trata de una estipulación terminológica que no coincide con otro uso habitual del término “reduccionismo”: muchas veces, se insiste en

que las formas más habituales de funcionalismo no son reduccionistas en la medida en que *no suponen la reductibilidad de los tipos funcionales mismos a tipologías más básicas*. Según mi estipulación terminológica, el funcionalismo analítico es normalmente reductivo: *trata de analizar los predicados mentales en términos no psicológicos*. Otra característica fundamental es su compatibilidad con una teoría materialista de la mente: el funcionalismo analítico (por ejemplo, del recientemente fallecido D. Lewis) puede identificar un estado mental de un sujeto o de una población con un tipo de estado material. Y puede hacerlo sin necesidad de renunciar a su pretensión de ser una teoría a priori: está establecido a priori que un estado (quizás desconocido) identificado por cierto rol funcional es el estado mental de un sujeto. Lo que no es a priori es cuál sea de hecho ese estado material. Su divergencia con las formas a posteriori de materialismo y de funcionalismo estriba en que, para éstas, la investigación empírica puede proporcionarnos criterios de identificación de estados mentales que ignoren y violen los criterios implícitos en las atribuciones cotidianas y accesibles al hablante ordinario. Como veremos, ésa no es una diferencia pequeña. En la aparición recalcitrante de ciertos problemas seculares (por ejemplo, el carácter supuestamente misterioso de la conciencia) es una diferencia crucial, que sitúa a los defensores de las dos formas de funcionalismo en territorios opuestos.

2. *Materialismo y psico-funcionalismo*

Actualmente, la mayoría de los filósofos analíticos de la mente se auto-clasificarían de *funcionalistas a posteriori*. Uno de los rasgos básicos de la discusión contemporánea no es sólo el carácter dominante de tal posición, sino el hecho de que, con ella, se ha tendido a borrar la distinción entre funcionalismo y materialismo que era común hace algunas décadas. En efecto, una vez se ha aceptado que la caracterización de las propiedades esenciales de un determinado tipo de estado mental depende de descubrimientos empíricos, que pueden imponer taxonomías y clasificaciones no isomorfas con las que son accesibles a priori a un hablante competente, parece artificioso insistir en la diferencia entre el rol causal y el realizador material. Después de todo, estados materiales idénticos en contextos idénticos son idénticos funcionalmente. Y cierta referencia mínima a un contexto causal parece necesaria si la tesis del *materialismo de la identidad* ha de ser plausible. El materialismo contemporáneo trata desesperadamente de no ser chauvinista: son concebibles seres dotados de mente con una composición material distinta a la nuestra. La identidad entre tipos de estados mentales y tipos materiales ha de ser concebida como relativa a una población determinada. Una vez aceptado este punto crucial, las diferencias con el funcionalismo a posteriori se reducen considerablemente: parece que cierta referencia al rol funcional, en sentido laxo, debe ser imprescindible para determinar si dos estados materiales diferentes tienen que contar o no como estados de tipos mentales semejantes cuando sean típicos en poblaciones muy diferentes desde el punto de vista material.

El *materialismo reductivo*, prevaleciente en la literatura filosófica actual sobre la mente, no se contrapone sólo al dualismo de sustancias: el dualismo cartesiano no es una de las opciones relevantes en la discusión contemporánea. El materialismo reductivo ha

de ser contrapuesto a formas no reductivas de materialismo. El *materialismo mínimo o no reductivo*, que comparten todas las posiciones relevantes en la filosofía de la mente actual, se caracteriza por aceptar el hecho de que las propiedades mentales sobrevienen a las propiedades básicas materiales. Dos mundos que fueran idénticos en las propiedades básicas de la materia y en sus relaciones no podrían diferir en otras propiedades: serían estética, económica y mentalmente iguales. No es este el momento de discutir el estatus de esa convicción. Una lectura atenta de la literatura nos convencerá fácilmente de que es defendida, en parte, por motivos conceptuales y, en parte, por motivos empíricos. Por motivos conceptuales: nuestra idea de lo que sea una propiedad básica es exactamente la idea de que pertenece a un tipo que genera esa relación de dependencia. Por motivos empíricos: tenemos buenas razones para creer que la metodología que sigue la física básica nos descubre propiedades cada vez más elementales. Sea como fuere, la tesis del materialismo reductivo no se reduce a la tesis del materialismo mínimo, necesita algo más: una tesis sobre la *dependencia sistemática* (un caso límite sería la identidad) entre los estados mentales y los estados materiales, más fuerte que la mera relación de *sobrevenida global*. Aceptar, por ejemplo, la tesis de la sobrevenida global de los hechos económicos sobre los estados físicos del universo no parece, al menos a primera vista, generar relaciones de dependencia metafísicamente interesantes entre el auge de la inflación y los hechos físicos. A primera vista, las distintas situaciones posibles que pueden describirse como situaciones de aumento desorbitado de los precios son situaciones que no tienen relaciones de semejanza interesantes desde el punto de vista de las propiedades físicas básicas involucradas. Desde este punto de vista, dos situaciones de inflación económica no tendrían que parecerse entre sí más de lo que podría parecerse cualquiera de ellas a una situación de deflación. Por supuesto, los defensores del materialismo reductivo piensan que tienen argumentos para mostrar que, al menos en el caso de los predicados mentales, sí hay razones para establecer tales relaciones de dependencia metafísica. Típicamente, identidad o, como mínimo, realización: un estado mental particular tiene los poderes causales que tiene en virtud de ejemplificar los poderes causales propios de ciertas propiedades básicas.

Entiendo, por tanto, que la división más profunda que recorre la filosofía de la mente actual no está bien caracterizada en términos de “realismo” versus “anti-realismo” ni en términos de “conductismo lógico” (o “funcionalismo analítico”) versus *materialismo o fisicalismo a posteriori*. La primera de las contraposiciones será normalmente rechazada por la mayoría de los que tendrían que ser catalogados como “anti-realistas”. La segunda de las contraposiciones olvidaría dos aspectos fundamentales de la discusión contemporánea: el carácter reductivo del conductismo lógico y el funcionalismo analítico y, por otra parte, el hecho de que éste último es perfectamente compatible con el materialismo a posteriori. La contraposición debe establecerse en términos de una cuestión crucial: *¿exige nuestra concepción científica del mundo que cualquier tipología —incluidas las mentales— que trate de atrapar propiedades genuinas, deba tener relaciones de dependencia sistemática y local con ciertos tipos de estados materiales o funcionales, que tendrían que ser determinados por la mejor ciencia posible?* Los defensores de las formas de materialismo y funcionalismo a posteriori, y también los funcionalistas analíticos que creen que su posición es compatible

con el materialismo responderían que sí. Un *eliminacionista* típico, también respondería que sí. Simplemente, rechazaría que haya genuinas propiedades mentales e insistiría en que la tipología impuesta por nuestros predicados mentales es completamente heteromorfa con la que la ciencia del cerebro nos revela. Es por ello por lo que evitaré cuidadosamente el término “*realismo intencional*” para referirme a la posición dominante en la filosofía de la mente actual y, en su lugar, utilizaré el término “reduccionismo” o “materialismo reductivo”. Por supuesto, se trata de una estipulación terminológica con el único propósito de individualizar un conjunto de teorías. Debemos recordar que un funcionalista puede defender que no trata de reducir la eficacia causal de los tipos funcionales a la eficacia causal de los tipos materiales más básicos. Es ésta una discusión sobre la que tendré que volver más adelante. En cualquier caso, mi justificación para la elección de la expresión ‘*materialismo reductivo*’ es que permite agrupar un conjunto de concepciones de la mente humana que comparten dos supuestos fundamentales:

- (a) la eficacia causal de un estado mental está determinada, en cada caso particular, por la eficacia causal de un estado material al que es idéntico o que constituye su realización en ese contexto y
- (b) los rasgos mentales son reductibles a propiedades materiales o funcionales.

3. Causalidad y Explicación psicológica

La sobreveniencia global de cualquier propiedad sobre las propiedades físicas no es un fenómeno decisivo a favor del reduccionismo. No es *prima facie* obvio, por ejemplo, que la sobreveniencia global de los hechos económicos, deportivos, estéticos o morales sobre los hechos físicos exija la reducción de unos a otros. Por razones expositivas y dialécticas, no voy a entrar en la forma especial en que la ausencia de reductibilidad se usa para fundamentar el eliminacionismo; pues la forma típica de reduccionismo materialista en filosofía de la mente no es eliminacionista. Por tanto, tiene razones para creer que la reducción ha de ser de hecho posible (no sólo que la reducción tendría que ser posible *si* los predicados mentales atraparán genuinas propiedades). Cualquier camino que vaya de la sobreveniencia global a la necesidad de la reducción ha de incorporar algunos supuestos expliquen por qué la opción de la reducción es más atractiva que la mera eliminación. Creo que es justo decir que los argumentos más socorridos, los argumentos que articulan el trasfondo sobre el que debe apoyarse el reduccionismo, dependen de una forma u otra de dos intuiciones fundamentales e íntimamente conectadas: *la eficacia causal de los estados mentales y el tipo particular de éxito que tienen las explicaciones psicológicas ordinarias*.

La idea de que la mente tiene eficacia causal sobre el mundo físico ha sido tradicionalmente uno de los supuestos básicos en el rechazo del dualismo y el epifenomenalismo. Pero la tendencia reduccionista que caracteriza a la filosofía contemporánea de la mente no se basa sólo en el hecho obvio de que hay relaciones causales entre mental y lo físico. No es injusto decir que se ha producido en las últimas décadas lo que puede describirse como una reacción anti-davidsoniana, combinada con la idea de

que debemos explicar satisfactoriamente el éxito de nuestras explicaciones psicológicas ordinarias. Davidson hizo famosa *una teoría de la identidad de casos* (cada suceso mental es también un suceso físico) que pretendía tener la virtud de ser compatible con una forma radical de no-reduccionismo: la identidad de casos era compatible con que no existiera ninguna relación de dependencia sistemática entre los tipos correspondientes. Las relaciones causales debían ser reducidas a relaciones entre sucesos físicos —de donde se seguía, según Davidson, que todos los sucesos causalmente efectivos, incluidos los mentales, debían ser también físicos. Por otra parte, los rasgos mentales que se mencionan en las explicaciones psicológicas ordinarias no se suponían reducibles a propiedades físicas del mundo. Es terreno común en la literatura contemporánea aceptar que una solución como la de Davidson, además de no incorporar, como vio Quine, un criterio no circular para la individualización de sucesos, comportaba la ineficacia causal de las propiedades mentales y, por lo tanto, eliminaba el rol propiamente explicativo de los rasgos mentales: un suceso mental no podría causar uno físico *en virtud de* ejemplificar algún rasgo mental. Por supuesto, desde ese punto de partida común, desde el rechazo del nominalismo davidsoniano, las rutas diferentes hacia el reduccionismo o hacia el anti-reduccionismo nacen de una diferencia fundamental. Un reduccionista pensará que el éxito de nuestras explicaciones psicológicas ordinarias necesita, dados los supuestos mínimos mencionados sobre la eficacia causal de la mente en el mundo físico, cierta relación de dependencia sistemática entre los tipos mentales y los tipos físicos. Un anti-reduccionista, lo negará. Y con ello se verá obligado — en la medida en que niegue la radical separación davidsoniana entre la metafísica de la causalidad y la epistemología de la explicación causal— a revisar algunos de los supuestos sobre la causalidad mental que a un reduccionista le parecen intocables. Por supuesto, dada la variedad de doctrinas que he agrupado bajo el genérico “reduccionismo,” no todos sus defensores están igualmente comprometidos con el valor de las explicaciones psicológicas ordinarias. Es compatible con el espíritu del reduccionismo la idea de que muchas de nuestras explicaciones ordinarias no atrapan verdaderas relaciones causales y que muchos de nuestros predicados mentales ordinarios no se verían reflejados en modo alguno en una ciencia psicológica madura. No es compatible con el reduccionismo, si es que quiere evitar el eliminacionismo, insistir en que la mente carece de eficacia causal o en que no puede haber explicaciones causales psicológicas satisfactorias.

Imaginemos la siguiente posibilidad: un genio laplaceano que tiene un conocimiento completo de las leyes básicas del universo, de las propiedades básicas y de su distribución. Un defensor del reduccionismo puede aceptar que ciertas capacidades epistemológicas nuestras podrían serle, en principio, ajenas. Nuestro genio podría no tener acceso a la clasificación del mundo en las clases de semejanza que nuestros predicados mentales establecen. Podría no tener predicados que atraparan algo semejante a nuestras categorías de creencia o intención, por ejemplo. Esa no sería, sin embargo, la cuestión crucial para un reduccionista. La cuestión crucial tendría que ser la de que, si creencias e intenciones son causalmente relevantes y nuestra apelación a ellas es explicativamente relevante, los tipos introducidos por nuestra categorización mental del

mundo *no* pueden ser completamente heteromorfos desde el punto de vista de las propiedades básicas del mundo que el genio laplaceano sí podría detectar. Como Fodor ha argumentado una y otra vez, si la explicación psicológica atrapara casos particulares bajo tipologías arbitrarias desde el punto de vista del verdadero orden causal del cosmos, su éxito y estabilidad serían tan misteriosos como pudieran serlo los de una posible psicología que estableciera relaciones de semejanza entre fenómenos según, por ejemplo, el día de la semana en el que sucedieran (Fodor 2000). O como ha expresado, desde un punto de vista más general, David Lewis:

La enorme mayoría de los aspectos sobrevenientes del mundo sólo son dados por disyunciones misceláneas infinitas de condiciones físicas infinitamente complejas. Por tanto, detectarlas, nombrarlas o pensar sobre una de ellas en un momento determinado es algo que está más allá de nuestro poder. Los rasgos mentales del mundo no están en absoluto más allá de nuestras capacidades epistémicas. Conglomerados finitos de partículas —nosotros— podemos seguir su rastro. Debe haber, por tanto, alguna suerte de simplicidad en ellos. (Lewis 1995, p. 415)

Es controvertido qué pueda seguirse de este argumento: para formas convencionales de funcionalismo, la conclusión —la simplicidad subyacente de los tipos mentales— es compatible con cierta autonomía explicativa, incluso causal, de los tipos funcionales respecto a los tipos neurofisiológicos. Para los protagonistas de la resurrección de las viejas teorías de la identidad de tipos, como, por ejemplo, Kim, la simplicidad exigible sólo puede ser garantizada por la identidad de los tipos mentales-funcionales con los tipos más básicos (J. Kim 1998, y 2005). Por otra parte, el rechazo de este argumento general es lo que caracteriza las posiciones que he descrito como anti-reduccionistas. Todas ellas han de rechazar como confuso el uso de la expresión “seguir su rastro” que aparece en la cita de D. Lewis. Parte del problema radica en que ese vocabulario parece exigir que la conducta de seguir el rastro (por ejemplo, la conducta a través de la cual se manifiesta la percepción de un estado mental en otro ser) habría de ser caracterizada de una manera neutra, independiente, del objeto que se supone rastreado. Ese es, en términos generales, el supuesto que rechazan las concepciones neo-wittgensteinianas de la mente. Posiciones a las que se ha aproximado recientemente uno de los que fue en su momento padre fundador del funcionalismo: H. Putnam. El problema con el argumento general que hemos mencionado, para estas concepciones, sería simplemente que sienta la cuestión por anticipado. El no-reduccionista dirá que el argumento de Lewis necesita de un supuesto discutible: el supuesto de que el estado mental de percibir intencionalidad en otros es un mero estado de detección, un estado caracterizable independientemente de cualquier vocabulario intencional. Por el contrario, y utilizando la terminología del propio Lewis, su adversario sostendrá que nosotros, conglomerados finitos de partículas, podemos formarnos estados mentales sobre estados mentales de otros seres (lo que Lewis denomina “seguir el rastro”) sin necesidad de postular que haya orden ni simplicidad (en términos de propiedades básicas) en los estados mentales atribuidos, precisamente porque tampoco hay orden ni simplicidad en el estado de percibirlos. Por supuesto, una posición semejante necesita de una justificación mucho más complicada de la que podría describirse en un comentario de este tipo. Tendría que mostrarle al reduccionista que la defensa esbozada del anti-reduccionismo no supone un compromiso con el eliminacionismo. ¿Por qué no equi-

vale a decir que la misma (supuesta) percepción de orden y simplicidad que está en la base de nuestra competencia ordinaria es, ella misma, una ilusión? En todo caso, ¿cómo se explica el hecho de que parezca que no es completamente ilusoria? Estas son cuestiones típicas que muestran que los supuestos ocultos en muchas de las discusiones contemporáneas en filosofía de la mente se remiten a zonas mucho más generales de la discusión filosófica. Parece que la única manera en que el anti-reduccionista puede defenderse es aceptando una concepción del lenguaje y la intencionalidad cercana a la defendida por el último Wittgenstein.

Sea como fuere, la literatura contemporánea parece encontrar dificultades para elaborar una noción coherente de causalidad mental que sea capaz de mostrar que el reto que plantea la cita de Lewis puede ser respondido satisfactoriamente. Uno de los supuestos fundamentales del funcionalismo es, por supuesto, que las propiedades mentales son causalmente eficaces y que, para serlo, su eficacia causal ha de estar sistemáticamente relacionada con la eficacia causal de las propiedades básicas del mundo. Otro es que ha de explicarse satisfactoriamente cierta autonomía de los tipos mentales respecto a los tipos estrictamente neurofisiológicos. La tensión parece evidente: la aceptación de tal autonomía mínima pone en cuestión el tipo requerido de perspicua dependencia sistemática entre los procesos causales involucrados. En los últimos años, J. Kim ha sido el más destacado defensor de la tesis de que el funcionalismo que no acepte una reducción de los tipos funcionales a los tipos materiales correspondientes es una actitud incoherente. La intuición básica del argumento de Kim parece clara: la supuesta autonomía explicativa de los procesos de orden superior (mentales, funcionales) no puede ser más que un fenómeno ilusorio si atendemos a ciertos requisitos plausibles sobre el funcionamiento de toda explicación genuinamente causal. So pena de negar el cierre causal del mundo físico o aceptar que los casos de causalidad mental son casos de sobredeterminación causal, parece necesario aceptar que la causa mental de un efecto físico ha de ser ella misma una causa física. Y si las causas mentales son causas físicas, las propiedades mentales son propiedades físicas. La supuesta autonomía explicativa de las tipologías mentales y funcionales debe descansar, en contra de las pretensiones habituales de los defensores del funcionalismo, en la identidad de las propiedades mentales y las propiedades básicas del mundo. En cierto sentido, los últimos años del siglo XX pueden describirse como una época de pérdida de la inocencia respecto a la viabilidad del modelo funcionalista en tanto que esencialmente distinto al materialismo de la identidad.¹⁴ Los más destacados intentos de bloquear el materialismo de la identidad se caracterizan por tratar de extraer lecturas metafísicas de ciertas asimetrías explicativas entre las propiedades básicas del mundo y las propiedades de orden superior. Stephen Yablo, por ejemplo, ha insistido en la última década en la proporcionalidad de la relación causal: las causas funcionales/mentales son más proporcionales a los efectos funcionales/intencionales, mantienen relaciones contrafácticas más robustas con ellos, que las supuestas causas neurofisiológicas o físicas (Yablo 2003). Tal tipo de teoría parece encajar perfectamente en una metafísica de las propiedades de acuerdo con la cual las propiedades son conjuntos de poderes causales y los poderes causales de las propieda-

des de orden superior (o determinables) son menos específicos que los poderes causales de las propiedades de orden inferior (o determinados), en el sentido en que éstas últimas son una manera especial de ejemplificar la propiedad de orden superior (Shoemaker 2001). Nadie duda de que, en cierto contexto explicativo, apelar a una propiedad de orden superior puede ser más relevante que apelar a una de orden inferior. Lo que se nos propone es que el fenómeno no afecta sólo a la relevancia pragmática o epistémica de la explicación causal: se trataría de un fenómeno metafísicamente relevante. Es discutible que tal tipo de movimiento no pueda ser acusado por un defensor de las teorías de la identidad de tipos de cometer una petición de principio. Ello estaría conectado con las dificultades para explicar la relación de multiple-realización como un caso especial de la relación metafísica entre determinable y determinado (Ehring 1996). Por otra parte, los defensores de posiciones radicalmente anti-reduccionistas pueden considerar esta situación dialéctica como una reducción al absurdo de los supuestos fundamentales del reduccionismo. De hecho, la misma noción de realización sería para ellos sospechosa: una suerte de engendro gramatical que poseería al mismo tiempo los rasgos de los universales y de los particulares (Steward 1997). Otra manera de expresar esta reacción sería la de decir que la teoría que se nos propone entraña la ausencia de criterios metafísicos de individuación del (supuestamente) único realizador, en cada caso particular, de la propiedad funcional (Putnam, 2000, y Corbí y Prades 2000).

4. *Externismo, contenido estricto, percepción e intencionalidad*

Hay puntos de fricción mucho más específicos entre una concepción reduccionista y una concepción anti-reduccionista de la mente. Un argumento anti-reduccionista típico consiste en apelar a ciertas intuiciones sobre la individuación del contenido que no parecen encajar fácilmente con el reduccionismo. Por ejemplo, el *externismo*, el hecho de que las relaciones efectivas con el medio sean constitutivas de la manera en la que individualizamos contenidos intencionales. Es cierto que, en principio, el reduccionismo puede combinarse con el externismo de maneras muy distintas. Un caso obvio es el de Fred Dretske, quien defiende una teoría externista incluso de los *qualia* perceptivos y una concepción reduccionista de la mente (Dretske 1995 y 1996). Otro caso lo constituirían las formas de funcionalismo de “largo alcance” que consideran que los procesos funcionales relevantes tienen que incluir ciertos rasgos del medio. Otra manera, típica, es la de introducir la noción de *contenido “estricto”* como la noción relevante en una explicación psicológica genuina. Lo que comparten estas concepciones es la necesidad de que esté metafísicamente determinado el componente interno del contenido. Que una ciencia psicológica madura tenga que apelar a contenidos estrictos o amplios, por ejemplo, no es tan relevante en este contexto como el supuesto de que, para explicar causalmente la conducta de un organismo, debe estar metafísicamente determinado el factor interno que es crucial en la explicación. Y no es fácil establecer los principios por los cuales tal determinación sería posible. Decir que dos individuos físicamente iguales comparten tal factor interno no es avanzar demasiado. Básicamente, porque el factor interno se supone que ha de ser compartido por individuos que no

sean exactamente iguales en todos los aspectos. El problema puede ser planteado en términos muy generales: una vez que aceptamos que nuestra manera ordinaria de clasificar contenidos mentales está comprometida con el contenido “amplio”, podemos clasificar conductas que son semejantes desde el punto de vista físico como formas muy distintas de acción intencional, y viceversa. Y ello hace mucho más verosímil uno de los puntos básicos de la resistencia del anti-reduccionista al argumento de Lewis que previamente citado: conductas muy diferentes desde el punto de vista físico pueden contar como el mismo tipo intencional, y conductas muy semejantes desde el punto de vista físico pueden contar como casos muy diferentes desde el punto de vista intencional. Necesitamos principios no *ad hoc* para individualizar los poderes causales que vayan más allá de la trivialidad de decir que siempre podemos conseguir bases de sobreveniencia suficientemente amplias tales que podemos asegurar que, dentro de ellas, se ejemplifica una determinada propiedad mental. Esa base acabaría siendo tan amplia que también garantizaría la ejemplificación de otras propiedades mentales independientes, por lo que no se habría establecido la base de sobreveniencia de esa propiedad mental particular que las estrategias reduccionistas requieren.

La idea de contenido estricto está vinculada a lo que tradicionalmente se ha denominado la “*Teoría Computacional de la Mente*”, la doctrina de que los poderes representacionales de la mente sólo pueden ser explicados si postulamos particulares mentales como portadores de tales poderes (representaciones) y explicamos las transacciones semánticas como el resultado de transacciones sintácticas, es decir, como interacciones causales entre diversas representaciones, en virtud de sus propiedades básicas y no representacionales. Ciertamente, un proceso computacional es a la vez semánticamente evaluable y explicable en virtud de los rasgos sintácticos, físicos, de los portadores de la representación. Pero un proceso computacional típico parece requerir la adjudicación arbitraria de contenido —en función de los intereses del diseñador y/o del usuario— a las representaciones básicas. No es de extrañar, pues, que la oposición típica a las teorías computacionales dependa de la intuición de que la apelación a procesos computacionales requeriría apelar a mecanismos básicos y no computacionales de fijación de contenido. El problema crucial es que no parece posible contar una historia creíble sin abrir la puerta a intuiciones externistas y, de ese modo, las concepciones computacionales de la percepción pierden su atractivo fundamental. Entre otras cosas porque, en ese caso, parece posible argumentar que es una ilusión suponer que la semántica de la mente *debe* estar necesariamente respaldada por clases de equivalencia desde el punto de vista de la neurofisiología. Sería perfectamente compatible con los datos empíricos el supuesto de que las clases de equivalencia intencionales fueran heteromorfas con clases de equivalencia más básica. La adjudicación de un contenido estable a partir de manifestaciones físicas bien distintas (y la adjudicación de contenidos distintos a manifestaciones físicas muy similares) podría ser explicada por el hecho de que la tipología intencional sólo es accesible desde una determinada actitud ante el mundo, ante la manera en que el agente atribuido se manifiesta en un mundo de objetos.

Por el contrario, los defensores de versiones más o menos fuertes de la teoría computacional insisten en que hay ciertos datos que serían inexplicables de otra manera: la composicionalidad y sistematicidad del lenguaje por ejemplo, o la infradeterminación del contenido perceptivo a partir de la escasez de los *inputs* causales. La reacción típica del anti-reduccionista es la de acusar al adversario de una descripción sesgada de los supuestos datos. La capacidad innata de descubrir regularidades en el lenguaje no requiere postular el tipo de representaciones internas que postula la teoría computacional. Lo mismo sucede con la supuesta infradeterminación de los estímulos perceptivos: después de todo, la teoría computacional depende de dos supuestos que están en tensión y que parecen difícilmente conciliables sin estipulaciones *previas* sobre los portadores primitivos de contenido. Por una parte, depende de la intuición galileana de que la percepción es una relación causal y de que, como en toda cadena causal, se da una dependencia contrafáctica mucho más robusta entre el contenido perceptivo y los estímulos próximos que la que se da entre el contenido perceptivo y los estímulos lejanos: el contenido-perceptivo-de-tigre es explicado mejor por la distribución de luz en la retina que por la presencia real y verdadera del tigre. Por otra parte, la teoría computacional necesita el supuesto de que los estímulos próximos no determinan el contenido: sin tal supuesto no puede describirse el fenómeno de la pretendida infradeterminación. Una manera anti-reduccionista típica de resolver la tensión es la de defender el realismo directo y una concepción disyuntiva del contenido perceptivo, según la cual el contenido perceptivo no es el mismo cuando el tigre está delante y cuando se produce la alucinación perfecta (McDowell 1994). De hecho, el movimiento equivale a negar la intuición galileana básica sobre la (relativa) independencia contrafáctica entre el contenido perceptivo y el estímulo externo. Y, desde ese punto de vista, las teorías computacionales tratarían de postular entidades *ad hoc* (representaciones) para salvar dos principios que son incoherentes. Sea cual sea el futuro desarrollo de esta controversia fundamental, es justo decir que las defensas más sofisticadas de las teorías computacionales son conscientes de algunas de sus dificultades: por ejemplo, el carácter local de los procesos sintácticos postulados sobre representaciones particulares parece difícil de encajar con el carácter holista de muchos procesos cognitivos fundamentales. Toda la controversia suscitada en relación a la relevancia filosófica de los modelos conexionistas está vinculada a esta cuestión fundamental (Fodor 2000b y Smolensky 1993). Por supuesto, el adversario de las teorías computacionales interpreta el problema como un subproducto de una confusión fundamental: no es que niegue a priori la existencia de procesos locales y causalmente relevantes para la génesis del contenido perceptivo. Niega el argumento por el cual sólo la existencia de relaciones causales entre los rasgos sintácticos de las representaciones puede resolver ciertas perplejidades supuestamente asociadas al éxito de procesos cognitivos básicos¹. El que los seres humanos representen el mundo no requiere que lo hagan por medio de representaciones internas. Para un reduccionista, su adversario parece condenado a negar la

¹ Una crítica exhaustiva a los supuestos fundamentales de las teorías computacionales de la mente puede encontrarse en Bennett y Hacker (2004).

intencionalidad, a afirmar que es sólo un asunto de interpretación de ciertas capacidades de nivel personal, del tipo de actitud que, contingentemente, adoptamos sobre la conducta abierta de nuestros semejantes. Para un anti-reduccionista, su adversario —al igual que los epistemólogos clásicos del XVII y del XVIII— es incapaz de ofrecer una explicación satisfactoria del hecho de que las supuestas representaciones internas adquieran el contenido intencional que se les supone. De nuevo, la discrepancia en este punto nos remite a un problema mucho más general: el del ajuste entre un sistema de conceptos y el mundo. Un reduccionista sólo puede mantener su acusación si insiste en que nuestras tipologías intencionales ordinarias —aquellas mediante las cuales “interpretamos” la conducta de nuestros semejantes— tienen una relación de ajuste con el mundo que puede ser descrita sin presuponerlas. Un anti-reduccionista niega que ese supuesto sea —en general, no sólo en relación con el vocabulario mental— inteligible y niega, además, que de su negación se siga que no hay hechos que conviertan en verdaderos o falsos nuestras atribuciones ordinarias de intencionalidad.

Para un reduccionista, la propuesta típica de su adversario —el “*realismo directo*” perceptivo— es una negativa a plantearse si siquiera las cuestiones fundamentales. Su adversario insiste en que el realismo directo no es una teoría alternativa: es un rechazo de los problemas mismos que se supone que han de ser respondidos por las teorías que defienden variantes más o menos sofisticadas de la vieja idea de los intermediarios. Como anteriormente comentábamos, el reduccionismo necesita suponer que hay una explicación razonable de la estabilidad mínima y la capacidad de predicción en la psicología de sentido común. En cierto sentido, y paradójicamente, el anti-reduccionista tiene un punto de coincidencia con él en este aspecto. Un punto sobre el que trata de alterar los términos de la discusión. El hecho de que nuestra percepción de la mentalidad de otros no sea sistemáticamente falsa es, para el anti-reduccionista, un caso particular de un fenómeno mucho más general que, según él, su adversario ignora sistemáticamente: el carácter genuinamente explicativo y exitoso (“*factive*”) de la noción de percepción, que es más básica que una noción de contenido perceptivo supuestamente neutra respecto al éxito (verdad) de la percepción². Y, en general, la idea de que la atribución de creencias está regida por el principio de maximización de la verdad y la racionalidad. El anti-reduccionista tenderá a considerar que ese principio, como un principio metodológico o epistémico que rige la atribución de creencias y, en general, de intencionalidad, puede servir para legitimar su negativa a entender la propuesta de su adversario. Se sigue de tal principio una asimetría epistémica fundamental entre las bases sobre las que atribuimos estados mentales y las bases sobre las que clasificamos, por ejemplo, los procesos causales físicos. Ése era el fundamento de la distinción radical propuesta por Davidson entre la metafísica de la mente y su epistemología. Las formas actuales de anti-reduccionismo, lo he mencionado, no siguen en eso a Davidson. Insisten, en cambio, en ciertas consecuencias metafísicas de tal diferencia

² Una defensa de la tesis de que los estados que incorporan conocimiento (conocer, percibir...) son psicológicamente más explicativos que sus correlatos “internos” y no necesariamente exitosos (creer, tener un determinado contenido perceptivo...) puede encontrarse en Williamson (2000).

epistémica. Sería injusto considerar que, al hacer tal cosa, están sentando por anticipado la cuestión contra su adversario reduccionista. Al fin y al cabo, el reduccionista parte, como hemos visto, de la necesidad de explicar el éxito (epistemológico) de nuestras atribuciones ordinarias de mentalidad. Considera, también, que el éxito epistemológico debe ser, en último término, el mejor valedor de la metafísica reduccionista que propone. Insiste, como vimos en la cita de Lewis, en que nuestra capacidad de detectar con éxito ciertos patrones en la conducta de nuestros semejantes es el mejor argumento a favor de la existencia de cierta relación de dependencia metafísica entre los tipos mentales y los tipos físicos.

De hecho, muchos de los supuestos subyacentes en las teorías computacionales de la percepción pueden considerarse una sofisticación de lo que desde Galileo se considera como la imagen científica mínimamente ortodoxa. La que se supone que se sigue del hecho de que la percepción es una relación causal ordinaria. La percepción se conceptualizó, por Galileo y Locke, básicamente en términos causales, y se consideró como un hecho obvio el que, en el último eslabón de la cadena causal, se produce un contenido que representa el mundo. Nunca nos dicen —como tampoco lo hacen los otros grandes epistemólogos del XVII y del XVIII— cómo es posible formarse contenidos perceptivos que se suponen sistemáticamente falsos (vgr. sobre colores) y que, por tanto, no pueden ser explicados en términos de las relaciones causales efectivas. Las dificultades reconocidas por todos —defensores y detractores— en las teorías computacionales de la percepción guardan un extraordinario parecido con los problemas asociados al representacionalismo clásico. La percepción es una relación causal. Por una parte, el contenido de la percepción está fijado por los últimos eslabones de la cadena causal correspondiente, como opuestos al estímulo remoto. Por otra, el efecto último de la cadena causal es, en el caso de la percepción, una representación. La tensión entre ambos supuestos resulta obvia en la manera en que se identifica el problema básico de la filosofía de la percepción: el problema de explicar cómo los últimos eslabones externos de la cadena causal —por ejemplo, los estímulos luminosos que llegan a la superficie de la retina— son procesados por el cerebro para dar lugar a un contenido que parece ir más allá de la información que está intrínsecamente en ellos. Aunque ya no se concibe que la función básica de los procesos internos de procesamiento sea la de justificar nuestra creencia perceptiva, la deuda a pagar sobre la génesis del contenido intencional en el caso de la percepción es todavía enorme. Antes de contar una historia inteligible sobre cómo es posible reconstruir la riqueza del estímulo externo a partir de las características intrínsecas del estímulo cercano (por ejemplo, la distribución de luz en la retina), necesitamos una historia creíble que nos explique cómo pueda haber contenido a procesar en las características intrínsecas de los estímulos: sin ese supuesto, no hay contraposición posible entre estímulos pobres en contenido y contenido perceptivo exuberante. No parece fácil contar una historia semejante sin

conceder al adversario algo fundamental: ciertas intuiciones externistas sobre los procesos que fijan el contenido³.

5. *La conciencia, de nuevo...*

La idea de que el modelo dominante sobre la percepción convierte a la intencionalidad en un misterio puede verse reforzada por el resurgimiento de una controversia que parecía que había perdido la virulencia que tuvo hasta mediados del siglo XX: el problema de la conciencia. La percepción es crucial porque los estados perceptivos tienen a la vez contenido intencional (representan el mundo) y aspectos fenomenológicos (son un caso de sentirse de cierta manera). Locke, por ejemplo, aceptó sin ambages que su concepción de la percepción convertía la aparición de los aspectos fenomenológicos de la misma en un verdadero misterio: las leyes científicas sobre la percepción que la mejor ciencia pudiera nunca descubrir, serían compatibles con la ausencia de aspectos fenomenológicos o con su distribución completamente distinta en el mundo material. Locke negaba que pudiéramos explicar alguna vez la supuesta sobreveniencia de la fenomenología de la percepción sobre los hechos físicos. De un modo similar, en su *Naming and Necessity*, S. Kripke planteó una objeción fundamental a lo que en este comentario se ha denominado reduccionismo a posteriori (Kripke 1980). Básicamente, el argumento se basaba en la imposibilidad de describir ciertos posibles descubrimientos empíricos como descubrimientos de la esencia oculta de los estados fenomenológicos. Kripke pareció suponer que todo descubrimiento a posteriori de la esencia oculta debería presuponer una diferencia entre los rasgos epistémicamente accesibles del fenómeno —aquellos a los que accede cualquier hablante competente— y los rasgos que constituirían su esencia oculta. En el caso del agua, el carácter de ser líquido, incoloro, inodoro etc., por contraposición a la composición química (H₂O). Esa diferencia debería generar la posibilidad de algo que pareciera normalmente acuoso sin ser agua, y al revés. En el caso de los estados fenomenológicos, definidos en términos de cómo se siente el sujeto, es ésa la diferencia que sería ininteligible. La década de los 90 del siglo pasado fue testigo de un resurgimiento de la disputa. Los argumentos neo-kripkeanos tenían la consecuencia de desbaratar la posibilidad misma de una teoría coherentemente materialista y a posteriori de la conciencia: tal teoría necesitaría establecer las condiciones de identidad metafísica de los estados de conciencia a partir de ciertos descubrimientos empíricos. Por una parte, se produjeron argumentos para tratar de mostrar que existían alternativas que el argumento original de Kripke no había tenido en cuenta. Alternativas basadas, por ejemplo, en la peculiaridad de los conceptos fenoménicos. La consecuencia del argumento de Kripke debería ser, según este tipo de estrategias, no la negación de las identidades a posteriori en el caso de la conciencia, sino la relevancia de la peculiaridad de los conceptos fenoménicos. Con ellos el sujeto se pone en una relación epistémica especial con ciertos estados neurofisiológicos que son

³ Para una visión panorámica sobre la relación entre las discusiones actuales en filosofía de la percepción y las discusiones de la epistemología tradicional, puede consultarse Noë y Thompson (2002).

—idénticos a— los estados fenomenológicos. A pesar de ello, los mismos estados que son representados por los conceptos fenoménicos pueden ser conocidos a través de otros medios de representación (los que se incluyen, por ejemplo, en una descripción neurofisiológica de los mismos) (Loar 1997, y Block y Stalnaker 1999). Este tipo de reacción ha sido considerada desde diversos puntos de vista como una maniobra *ad hoc*, para salvar los compromisos mínimos del materialismo a posteriori. Se ha argumentado, por ejemplo, que la solución propuesta viola claramente las condiciones de inteligibilidad de la noción misma de identidad a posteriori (basadas en que haya siempre una diferencia entre lo presentado y el modo de presentación) y que la apelación al carácter especial de los conceptos fenoménicos en nada ayuda a entender el vocabulario de la identidad en este caso (Chalmers 1996 y 2003, y White 2006).

La cuestión es crucial porque afecta a la inteligibilidad misma de la forma dominante de materialismo en la literatura actual. En su libro más reciente, el propio J. Kim reconoce que su modelo reduccionista debe dejar a un lado, por este tipo de motivos, los aspectos fenomenológicos de la conciencia (Kim 2005). Y este tipo de problemas están a la base de la resurrección de formas neo-cartesianas de “misterianismo”: la idea de que, en el fondo, no podemos entender la emergencia de los aspectos fenomenológicos de la conciencia a partir de ciertos complicados procesos neurofisiológicos (Levine 2001). Como ha advertido recientemente N. Block (un autor que se situaría claramente en el campo del materialismo a posteriori), el problema es mucho más grave: aun suponiendo que existiera una solución aceptable al problema que respetara las premisas básicas del reduccionismo a posteriori, el precio a pagar, impuesto por la *forma* que necesariamente debería tener la supuesta solución, sería extraordinario: nos dejaría inermes frente a los problemas epistemológicos que tradicionalmente se han venido asociando a la fenomenología —espectros invertidos, otras mentes (Block 2002). Una vez realizada la correspondiente identidad a posteriori, no tendríamos ningún motivo racional para asumir que seres con neurofisiología —o con una constitución interna— radicalmente distinta a la nuestra tienen conciencia. Y tal conclusión debería ser completamente independiente de las capacidades y disposiciones funcionales que les fueran justificadamente atribuibles desde la tercera persona. El precio parecería claramente excesivo. Y, sin embargo, las teorías a las que ese problema no se plantearía (formas de funcionalismo a priori, o lo que en este artículo he denominado “anti-reduccionismo”) no parecen gozar de un respaldo mayoritario en la controversia contemporánea. Todo lo contrario.

6. Significado y atribuciones de mentalidad

Terminaré trazando una última conexión entre las dos posiciones generales que he tratado de dibujar (reduccionismo *versus* anti-reduccionismo) y ciertas controversias básicas en filosofía del lenguaje. Es importante ver que el anti-reduccionista puede utilizar, y utiliza, en favor de su posición ciertas concepciones del lenguaje y la intencionalidad sobre las que se puede argumentar con independencia de su relevancia para el problema de la eficacia causal de la mente. Por ejemplo, una concepción wittgensteiniana de la estabilidad en el uso de nuestros predicados y nuestras atribuciones de mentalidad,

O, por mencionar discusiones más recientes, una concepción contextualista y/o pragmatista del significado y de las atribuciones de intencionalidad. La conexión con este tipo de discusiones es múltiple. Mencionaré sólo su aspecto más general. Una teoría reduccionista sobre la mente necesita ser aplicable a los estados representacionales y a los contenidos intencionales. El tipo de “realismo robusto” que defiende requiere que los poderes causales de la representación atribuida sean sistemáticamente dependientes de los poderes causales de la base física, o neurofisiológica, que constituye, o es idéntica, o realiza el estado representacional atribuido al atribuir a alguien la creencia de que mañana lloverá. Si hubiera algo de verdad en los proyectos contextualistas en semántica, tendría repercusiones sobre el tipo de estado que atribuyo cuando atribuyo a alguien tal creencia. Si la atribución de tal creencia está radicalmente afectada por factores contextuales y pragmáticos, sería difícil defender la existencia de ciertos poderes causales comunes y describibles en términos no intencionales que pudieran sustentar la mínima unidad causal de la representación atribuida tal y como la concibe el proyecto reduccionista. Es parte del proyecto contextualista en semántica apelar a la pertinencia de utilizar el mismo predicado o el mismo tipo de atribución con finalidades y supuestos muy distintos. La percepción de la semejanza relevante para el uso del mismo tipo de predicado es lo que nos hace competentes lingüísticamente. Pero, si el contextualismo semántico tuviera algo de razón, sería inevitable su extensión a las atribuciones de intencionalidad. Y, con ello, parecería fuera de lugar el supuesto de que nuestras representaciones —los estados intencionales que atribuimos al atribuir intencionalidad— tienen los rasgos que un defensor del reduccionismo necesita: el hecho de que un agente ejemplifique un estado intencional de determinado tipo es relativo al contexto dialéctico en que se produce la atribución. En este punto es importante advertir que el reduccionista no puede argumentar que, aunque el proyecto contextualista en semántica tuviera básicamente razón, ello no implica que no haya verdaderos estados representacionales que no son atrapados por nuestras atribuciones ordinarias y que son los que están afectados por la tesis general del reduccionismo. El reduccionista necesita aquí, como lo necesita en el caso del externismo de lo mental, una historia creíble sobre la manera en que tales representaciones metafísicamente genuinas están relacionadas con nuestras atribuciones ordinarias; pues, recordémoslo, el tipo de reduccionismo contemporáneo que estoy discutiendo es profundamente antidavidsoniano: asume que una parte fundamental de su proyecto es el de hacer inteligible la estabilidad y el éxito de nuestras explicaciones psicológicas ordinarias⁴.

REFERENCIAS

- Alter, T. y S. Walter (eds.) (2006). *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.
- Bennett, M., y P. Hacker (2003). *The Philosophical Foundations of Neuroscience*. Malden, MA: Blackwell.

⁴ El ataque más radical en la literatura contemporánea a la distinción tradicional entre semántica y pragmática, aplicado a la metafísica del contenido intencional, se encuentra en Travis (2000).

- Block, N. (2002). "The harder problem of consciousness", *Journal of Philosophy* 99, pp. 391-425.
- , O. Flanagan y G. Güzeldere (eds.) (1997). *The Nature of Consciousness: Philosophical Debates*. Cambridge MA: MIT Press.
- , y R. Stalnaker (1999). "Conceptual analysis, dualism, and the explanatory gap", *Philosophical Review* 108, pp. 1-46.
- Corbí, J.E., y J.L. Prades (2000). *Minds, Causes, and Mechanisms. A Case against Physicalism*. Oxford: Blackwell Publishers.
- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- (2003). "Consciousness and its place in nature", en Stich y Warfield, *Blackwell Guide to Philosophy of Mind*. Oxford: Blackwell.
- Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.
- (1996). "Phenomenal externalism or if meanings ain't in the head, where are qualia?", *Philosophical Issues* 7, pp. 143-158.
- Fodor, J. (1998). *In Critical Condition*. Cambridge, MA: MIT Press.
- (2000a). "A Science of Tuesdays", *London Review of Books* 22.
- (2000b). *The Mind doesn't work that way*. Cambridge, MA: MIT Press.
- Gillet, C., y B. Lower (2001). *Physicalism and Its Discontents*. Cambridge: Cambridge University Press.
- Guttenplan, S. (1995). *A Companion to the Philosophy of Mind*. Oxford: Blackwell.
- Lewis, D. (1995). "David Lewis: Reduction of Mind", en S. Guttenplan, *A Companion to the Philosophy of Mind*. Oxford: Blackwell.
- Kim, J. (1998). *Mind in a Physical World: an essay on the mind-body problem and mental causation*. Cambridge, MA: MIT Press.
- (2005). *Physicalism, or something near enough*. Princeton: Princeton University Press.
- Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Levine, J. (2001). *Purple Haze: The Puzzle of Consciousness*. Oxford: University Press.
- Loar, B. (1997). "Phenomenal states", en N. Block, O. Flanagan y G. Güzeldere (eds.), *The Nature of Consciousness: Philosophical Debates*. Cambridge MA: MIT Press.
- McDowell, J. (1994). "The Content of Perceptual Experience", *Philosophical Quarterly* 44, pp. 190-205.
- Noë, A., y E. Thompson (eds.) (2002). *Vision and Mind: Selected Readings in the Philosophy of Perception*. Cambridge, MA: MIT Press.
- Putnam, H. (2000). *The Threefold Cord*. New York: Columbia University Press.
- Shoemaker, S. (2001). "Realization and Mental Causation", en Gillet and Lower, *Physicalism and Its Discontents*. Cambridge: Cambridge University Press.
- Smolensky, P. (1993). *On the proper treatment of connectionism*. Cambridge, MA: MIT Press.
- Steward, H. (1996). *The Ontology of Mind*. Clarendon Press: Oxford.
- Stich, P., y T. Warfield. (2003). *Blackwell Guide to Philosophy of Mind*. Blackwell: Oxford.
- Travis, C. (2000). *Unshadowed Thought: Representation in Thought and Language*. Cambridge, MA: Harvard University Press.
- White, S. (2006). "Property dualism, phenomenal concepts, and the semantic premise", en T. Alter y S. Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.
- Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.
- Yablo, S. (2003). "Causal Relevance", *Philosophical Issues* 13, pp. 316-329.

Josep L. PRADES es profesor titular de Lógica y Filosofía de la Ciencia en la Universidad de Girona. Anteriormente, fue profesor en la Universidad de Murcia. Sus publicaciones incluyen trabajos sobre la filosofía del segundo Wittgenstein, Epistemología y Filosofía de la Mente. Es autor, junto con Josep Corbí, del libro *Minds, Causes and Mechanisms*, Blackwell, Oxford, 2000. Actualmente, está interesado en problemas relacionados con la teoría de la acción y las atribuciones de intencionalidad.

DIRECCIÓN: Departament de Filologia i Filosofia. Facultat de Lletres. Universitat de Girona. Pl. Ferrater Mora, 1. 17071 Girona. E-mail: josepll.prades@udg.es.