

EL PROBLEMA MENTE-CUERPO TRAS CINCUENTA AÑOS

The mind-body problem after fifty years

Jaegwon KIM*

Brown University (USA)

BIBLID [(0213-356) 4, 2002, 45-63]

RESUMEN

Hace aproximadamente 50 años que se reintrodujo el problema mente-cuerpo en la filosofía como problema metafísico serio. Este artículo revisa el debate que ha seguido a las obras seminales de los años 50 y 60 de escritores tales como J. J. C. Smart, Herbert Feigl, Hilary Putnam y otros, y ofrece una evaluación del estado actual de la discusión.

Palabras clave: el problema mente-cuerpo, fisicalismo, reducción, supervenencia, emergencia, causación mental, conciencia.

ABSTRACT

It has been roughly 50 years since the mind-body problem was reintroduced into philosophy as a serious metaphysical problem. This paper reviews the debate that has followed the seminal works in the '50s and '60s by such writers as J. J. C. Smart, Herbert Feigl, Hilary Putnam, and others, and offers an assessment of the current state of the discussion.

Key words: the mind-body problem, physicalism, reduction, supervenience, emergence, mental causation, consciousness.

* Este artículo es una traducción de «The mind-body problem after fifty years», en: Anthony O'Hear (ed.), *Current Issues in Philosophy of Mind*, Cambridge University Press, Cambridge (UK), 1988, 3-21.

Agradecemos al autor su permiso para publicar en este número el artículo en castellano. Traducido del inglés por Jesús Vega Encabo y José Luis Vega Encabo.

I

Fue hace medio siglo cuando el problema mente-cuerpo, moribundo como muchos otros en la metafísica sería durante varias décadas, se resucitó como problema filosófico de primera fila. El primer impulso procedió de *The Concept of Mind* de Gilbert Ryle (publicado en 1948) y de las reflexiones, extensamente conocidas, aunque no bien entendidas, sobre la naturaleza de lo mental y el lenguaje mental que hizo Wittgenstein en su *Philosophical Investigations* (publicadas en 1953). Sin embargo, las inquietudes principales, tanto de Ryle como de Wittgenstein, se enfocaron más sobre la lógica del discurso mental que sobre la cuestión metafísica de cómo se relaciona nuestra mente con nuestra naturaleza corporal. De hecho, Ryle y Wittgenstein, cada uno por diferentes razones, habrían considerado que el problema metafísico de la relación mente-cuerpo se debía a deplorables confusiones lingüísticas y que no era abordable mediante una discusión inteligible. Por aquella época, ya circulaba el anterior y ampliamente ignorado clásico *The Mind and Its Place in Nature* de C. D. Broad, aparecido en 1925. Sin embargo, esta obra, aunque robustamente metafísica, no logró conectarse con (y configurar) el debate mente-cuerpo en la segunda mitad de ese siglo. En verdad, el problema mente-cuerpo tal como lo conocemos hoy en día tuvo sus orígenes cercanos en un trío de artículos publicados a finales de los 50: «Is Consciousness a Brain Process?» de U. T. Place¹ en 1956, «Sensations and Brain Processes» de J. J. C. Smart, y «The “Mental” and the “Physical”», de Herbert Feigl, publicados respectivamente en 1958 y en 1959². En estos artículos, Place, Smart y Feigl proponían un acercamiento al estatuto de la mente que se ha llamado de varias maneras: «materialismo de la identidad mente-cuerpo», «materialismo del estado central», «fiscalismo de tipos» y «teoría del estado cerebral». Fueron principalmente los artículos de Feigl y Smart los que tuvieron un mayor impacto filosófico, pues impulsaron el debate que se ha continuado hasta hoy en día.

Para los que crecimos filosóficamente en la época de los 60, la teoría del estado cerebral fue nuestro primer encuentro con el problema mente-cuerpo como

1. U. T. PLACE, «Is Consciousness a Brain Process?», *British Journal of Psychology*, 47/1 (1956), pp. 44-50. Hubo incluso formulaciones modernas más tempranas del enfoque de la identidad: por ejemplo, Samuel ALEXANDER, *Space, Time, and Deity*, vol. II, Londres, McMillan, 1920, p. 9, donde dice: «Los procesos mentales y sus procesos neuronales son una y la misma existencia, no dos existencias»; el psicólogo Edwin G. BORING afirma: «si llegáramos a encontrar una correlación perfecta entre la sensación A y el proceso neuronal a, una correlación precisa de la cual tuviéramos razones para creer que nunca falla, deberíamos entonces identificar A y a... es científicamente más apropiado considerar que todos los datos psicológicos del mismo tipo y que la conciencia es un acontecimiento fisiológico», (*The Physical Dimensions of Consciousness*, New York, Dover, reimpresión de 1963, p. 14). El libro de Boring fue publicado por primera vez en 1933.

2. J. J. C. SMART, «Sensations and Brain Processes», *Philosophical Review*, 68 (1959), pp. 141-156. Herbert FEIGL, «The “Mental” and the “Physical”», en Herbert FEIGL, GROVER MAXWELL y MICHAEL SCRIVEN (eds.), *Minnesota Studies in the Philosophy of Science*, vol. II, Minneapolis, University of Minnesota Press, 1958.

problema dentro de la filosofía sistemática. Nos impresionó su refrescante audacia y parecía estar en consonancia con el temperamento científico tan optimista de aquella época. ¿Por qué no puede lo mental resultar ser procesos cerebrales tal y como el calor resultó ser movimiento molecular y la luz ondas electromagnéticas? Sin embargo, la teoría del estado cerebral tuvo una vida sorprendentemente corta; su precipitado declive empezó sólo unos pocos años más tarde de su promulgación inicial. Alrededor de finales de los 60 y principios de los 70 había sido abandonada por casi todos los filósofos que trabajaban en filosofía de la mente y psicología. Esto significó algo más que el desvanecimiento de una prometedora y arriesgada teoría filosófica: la muerte de la teoría del estado cerebral dio mal nombre a todas las formas de reduccionismo y convirtió el término «reduccionista» en un epíteto claramente negativo y, a menudo, desdeñoso. Llamar a alguien reduccionista hoy en día ha llegado a ser, en la mayoría de los círculos intelectuales, tanto como decir que mantiene una visión incorrecta; se trata de una humillación finamente camuflada que tacha a esa persona de intelectualmente reaccionaria y simplista.

No obstante, retrospectivamente, está claro que, a pesar de su corta vida, el fisicalismo de Smart-Feigl dejó una contribución crucial que ha sobrevivido a su reinado como teoría de la mente; lo que quiero decir es que esta teoría ayudó a sentar los parámetros básicos para los debates que habrían de seguirse después; un conjunto de presupuestos y aspiraciones ampliamente fisicalistas que todavía hoy guían y constriñen nuestro pensamiento. Un indicio de ello es que, cuando la teoría del estado cerebral se derrumbó, los filósofos no regresaron al cartesianismo ni a ninguna otra forma seria de dualismo. Casi todos los participantes en este debate se mantuvieron dentro del marco fisicalista, e incluso aquellos que tomaron parte principalmente en el fallecimiento del materialismo de Smart-Feigl continuaron siendo fieles a la visión fisicalista. Esto ha desempeñado un papel central en la definición de nuestra *Problematik*: a lo largo de los setenta y ochenta hasta nuestros días, el problema mente-cuerpo (nuestro problema mente-cuerpo) ha sido el de encontrar un lugar para la mente en un mundo fundamental y esencialmente físico. Si C. D. Broad escribiera ahora el libro que escribió en 1925 le podría haber dado perfectamente el título de *The Mind and Its Place in the Physical World*.

Lo que produjo el fallecimiento tan rápido y aparentemente sin dolor de la teoría del estado cerebral, causando pocos lamentos entre los filósofos, fue que la objeción principal que anunció su muerte, el denominado argumento de la realización múltiple (o «variable», como se dice en el Reino Unido), introducido en primer lugar por Hilary Putnam³, contenía dentro de sí semillas de un atractivo enfoque alternativo, a saber, el funcionalismo. La tesis central del funcionalismo (que las clases mentales son «clases funcionales», no físicas ni biológicas) era una

3. En «Psychological Predicates», publicado por primera vez en 1968 y reimpresso más tarde con un nuevo título, «The Nature of Mental States», en Hilary PUTNAM, *Collected Papers II*, Cambridge, Cambridge University Press, 1975.

idea atractiva y muy reveladora que parecía ayudarnos a dar sentido a una «ciencia cognitiva» que estaba siendo creada por entonces. La concepción funcionalista de la mente parecía hecha a medida para la nueva ciencia de lo mental y de la cognición, porque parecía postular un dominio autónomo y distintivo para las propiedades mentales/cognitivas que se podían investigar científica e independientemente de las encarnaciones físico/biológicas, una idea que prometía tanto legitimidad como autonomía para la psicología en cuanto ciencia. El funcionalismo nos permitió despojarnos de las limitaciones restrictivas del reduccionismo fisicalista sin regresar a los desacreditados dualismos de Descartes y otros. Al menos, esa impresión daba en la época. El concepto funcionalista de lo mental representa aún «la historia oficial» sobre la naturaleza y la fundación de la ciencia cognitiva⁴.

Sin embargo, los funcionalistas, por lo general, no eran metafísicos, y muy pocos se preocuparon siquiera de modo autoconsciente sobre el lugar en que se situaba el funcionalismo respecto al problema mente-cuerpo. Algunos funcionalistas, como David Armstrong y David Lewis, pensaban que defendían el fisicalismo, mientras que otros, como Hilary Putnam y Jerry Fodor, afirmaban que el funcionalismo proporcionaba una refutación decisiva del fisicalismo. El término clave que utilizaban para describir la relación entre las propiedades mentales (clases, estados, etc.), y las físicas era «realización» (o algunas veces «implementación», «ejecución», etc.): las propiedades mentales son «realizadas» o «implementadas» por (o en) las propiedades físicas, aunque no son ni idénticas ni reducibles a ellas. El término «realización» fue introducido⁵ y rápidamente ganó vigencia, principalmente en base a analogías computacionales (concretamente, máquinas de computación caracterizadas matemáticamente que se realizaban en ordenadores físicos), pero pocos funcionalistas, especialmente en los primeros días, se esforzaron por explicar en qué consistía la relación de realización, qué implicaba esta relación respecto a las opciones tradicionales sobre el problema mente-cuerpo.

Creo que la idea de «sobreveniencia» empezó a destacar en los 70 y 80 en parte para colmar este vacío. La doctrina de que las propiedades mentales sobrevienen a las propiedades físicas parecía responder bien a las necesidades de los fisicalistas postreduccionistas que buscaban una metafísica de la mente, ya que prometía un sentido claro y vigoroso a la primacía del dominio físico y sus leyes, y vindicaba así los compromisos fisicalistas de la mayor parte de los funcionalistas. Al mismo tiempo, los liberaba de las cargas del reduccionismo físico y, con ello, protegía lo mental como un dominio autónomo. Es más, al permitir múltiples bases físicas para las propiedades mentales sobrevinientes podía también acomodar la múltiple realizabilidad de las propiedades mentales. Muchos filósofos, especialmente aquellos que por una

4. Véase, por ejemplo, Zenon PYLYSHYN, *Computation and Cognition*, Cambridge, Ma., MIT Press, 1985.

5. El primer uso filosófico, que yo sepa, de este término, más o menos en su sentido actual, aparece en Hilary PUTNAM, «Minds and Machines», en Sydney HOOK (ed.), *Dimensions of Mind*, New York, New York University Press, 1960.

u otra razón habían abandonado toda esperanza de reducción fiscalista de lo mental, buscaron en la sobrevenida mente-cuerpo una formulación metafísica satisfactoria del fiscalismo sin reduccionismo. Al final de los setenta, lo que Ned Block ha llamado oportunamente «el consenso antireduccionista»⁶, estaba ya firmemente establecido. Esto ha ayudado a entronizar al «fiscalismo no reductivo» como la nueva ortodoxia, no sólo sobre la relación mente-cuerpo, sino, de manera más general, sobre la relación entre propiedades «superiores» y propiedades subyacentes «inferiores» también en todos los otros dominios. De este modo, el enfoque producía como prima una visión general fundamentada sobre la relación entre las ciencias especiales y la física básica.

Un efecto colateral del atrincheramiento del consenso antirreduccionista ha sido el retorno del emergentismo, quizá no la doctrina hecha y derecha del emergentismo clásico de los 20 y los 30, pero sí al menos de su vocabulario y consignas características. Cuando prevalecían el positivismo y la idea de «unidad de la ciencia», el emergentismo fue frecuentemente considerado con abierta desconfianza, como una doctrina metafísica misteriosa y posiblemente incoherente. Con el fiscalismo reduccionista caído en desgracia, el emergentismo se rehabilita con fuerza⁷, y se percibe ahora un incremento y un uso sin arrepentimiento de términos como «emergencia», «característica emergente», «fenómeno emergente», «causa emergente» y otros parecidos, más o menos en el sentido de los emergentistas clásicos, no sólo en importantes escritos filosóficos⁸ sino también en la literatura científica primaria de muchos campos⁹.

Resumiendo entonces, hay tres ideas que han destacado en el escenario de los recientes debates sobre la relación mente-cuerpo: la idea de que lo mental es «realizado» por lo físico, la idea de que lo mental «sobreviene» a lo físico, y la idea de que lo mental es «emergente» a partir de lo físico. En este artículo quiero explorar la interacción entre estas tres ideas y los papeles que desempeñan en los actuales debates sobre el problema mente-cuerpo, y con ello deseo indicar dónde considero que nos encontramos ahora en relación con este problema.

6. En su «Antirreductionism Slaps Back», en *Philosophical Perspectives*, 1997.

7. Además de un número de títulos recientes de revista, los signos del retorno del emergentismo incluyen una colección reciente de nuevos ensayos sobre la emergencia, *Emergence or Reduction?* A. BECKERMANN, H. FLOHR y J. KIM, (eds.), Berlin, de Gruyter, 1992, dos volúmenes de ensayos sobre la emergencia que se preparan en Europa como esta obra, y el Oberlin Philosophy Colloquium de 1997 sobre el tópico «Reduccionismo y emergencia».

8. Véase por ejemplo John R. SEARLE, *The Rediscovery of Mind*, Cambridge, Ma., MIT Press, 1992.

9. Por ejemplo FRANCISCO VARELA, EVAN THOMPSON y ELEANOR ROSCH, *The Embodied Mind*, Cambridge, Ma., MIT Press, 1991. Véase especialmente la Parte IV, titulada «Variedades de emergencia».

II

Empecemos por la sobreveniencia. Conviene interpretar la sobreveniencia como una relación entre dos conjuntos de propiedades, las propiedades sobrevenientes y sus propiedades «base». Como es bien sabido, hay disponible una gran variedad de relaciones de sobreveniencia, pero para nuestros propósitos actuales no se requieren distinciones meticulosas. La idea esencial de la sobreveniencia mente-cuerpo es que las propiedades o estados mentales de algo son dependientes de sus propiedades físicas o corporales, en el sentido de que, una vez que se fijan sus propiedades físicas, se fijan con ello sus propiedades mentales. Esto implica que si dos cosas –organismos, personas o sistemas electromecánicos– tienen propiedades físicas idénticas deben tener también naturalezas mentales idénticas; es decir, gemelos físicos exactos son *ipso facto* gemelos mentales exactos. La sobreveniencia mente-cuerpo se puede útilmente formular así de manera equivalente: si un organismo instancia una propiedad psicológica *M* (digamos, dolor) en un momento dado, entonces tiene en ese momento alguna propiedad física *P* sobre la que *M* sobreviene, tal que necesariamente si algo tiene *P*, tiene *M*. Así, si se experimenta dolor, se debe instanciar en ese momento una cierta propiedad física (presumiblemente alguna propiedad neuronal) sobre la que sobreviene el dolor. Ninguna propiedad mental puede ser instanciada en un organismo a menos que ese organismo instancie alguna propiedad física apropiada que sirva como su base física.

De esta manera, la sobreveniencia mente-cuerpo prometía dar sentido a la idea fisicalista de que lo físico goza de primacía ontológica frente a lo mental, y a la idea de que la física es la más básica y más comprehensiva de nuestras ciencias, siendo todas las restantes «ciencias especiales» sobre dominios restringidos. Además, y esto era de crucial importancia para el fisicalista no-reductivo, la sobreveniencia no parecía *prima facie* comprometernos con el reduccionismo: después de todo, muchos teóricos morales, como G. E. Moore y R. M. Hare, creían en la sobreveniencia de lo moral sobre lo no moral, pero rechazaban la reducibilidad de lo primero a lo segundo. Además, aceptar la idea de que las propiedades estéticas de las obras de arte son sobrevenientes sobre sus características físicas no parece conducir a la posición de que las propiedades estéticas sean reducibles a propiedades físicas. Parecía, entonces, que finalmente habíamos podido encontrar en la sobreveniencia mente-cuerpo una base metafísica para el fisicalismo no reductivo; la sobreveniencia parecía ser precisamente la relación metafísica de dependencia que nos permitiría comprender cómo lo mental, a pesar de su dependencia de lo físico, podía aún seguir siendo irreducible a él y formar su propio dominio autónomo.

Gran parte de la discusión que siguió a la introducción de la idea de sobreveniencia en el debate mente-cuerpo versó sobre la cuestión de si la sobreveniencia estaba efectivamente libre de implicaciones reduccionistas. La cuestión está todavía pendiente, pero ha quedado claro que en realidad no era el asunto. El auténtico asunto, creo, es si la doctrina de la sobreveniencia mente-cuerpo puede o no

puede en sí misma reivindicarse como una posición distintiva sobre el problema mente-cuerpo. La cuestión, entonces, es ésta: ¿tenemos en la sobrevenida mente-cuerpo una explicación de cómo se relaciona nuestra mente con la naturaleza física de nuestro ser? O lo que es lo mismo, ¿podemos usar la misma sobrevenida para exponer una teoría filosófica sobre cómo las mentes se relacionan con los cuerpos?

Una breve reflexión nos demuestra que la respuesta es no: la sobrevenida mente-cuerpo no puede por sí misma constituir una teoría de la relación mente-cuerpo. Existen para esto dos razones relacionadas entre sí. En primer lugar, la sobrevenida mente-cuerpo es consistente con una multitud de posiciones clásicas sobre el problema mente-cuerpo; de hecho es un compromiso compartido por muchas teorías mente-cuerpo mutuamente excluyentes. Consideremos, por ejemplo, el emergentismo: el emergentismo es una teoría dualista que enfatiza la irreductibilidad de los emergentes, incluyendo las propiedades mentales, a condiciones psicoquímicas más básicas, y, con todo, respeta la sobrevenida. En el emergentismo, los emergentes necesariamente emergen cuando y sólo cuando están presentes las «condiciones basales» apropiadas; cuando están presentes idénticas condiciones basales, idénticos emergentes deben emerger. La concepción funcionalista de que lo mental, cuando se realiza, debe tener también una realización física, implica la sobrevenida mente-cuerpo: para las mismas condiciones físicas, las mismas propiedades funcionales. Y lo que es más obvio, la sobrevenida mente-cuerpo es una consecuencia trivial del fisicalismo de tipos (por ejemplo, de la teoría del estado cerebral), que identifica reductivamente las propiedades mentales con las propiedades físicas. Incluso el epifenomenalismo está comprometido con la sobrevenida; si dos cosas difieren en algún aspecto mental, debe ser a causa de que difieren en algún aspecto físico; debe ser porque la causa física del aspecto mental involucrado está presente en uno y ausente en otro. Si la sobrevenida mente-cuerpo es un compromiso de cada uno de estos enfoques en conflicto al problema mente-cuerpo, no puede ser por sí misma una postura sobre este tema al lado de estas alternativas clásicas¹⁰.

Todo ello muestra que el mero hecho (asumiendo que sea un hecho) de la sobrevenida mente-cuerpo deja abierta la cuestión acerca de lo que la fundamenta y la explica, es decir, la pregunta acerca de por qué la relación de sobrevenida se da entre lo mental y lo físico¹¹. Para comprender el asunto general aquí implicado, considérese la sobrevenida normativa, la doctrina ampliamente aceptada de que

10. La sobrevenida mente-cuerpo no está excluida tampoco por el dualismo de sustancia cartesiano. Véase mi «Supervenience for Multiple Domains», reimpresso en *Supervenience and Mind*, Cambridge, Cambridge University Press, 1993.

11. Sobre la necesidad de explicar la relaciones de sobrevenida, véase Terence HORGAN, «Supervenience and Cosmic Hermeneutics», *Southern Journal of Philosophy*, 22 (1984), Suplemento, pp. 19-38, y Terence HORGAN y Mark TIMMONS, «Troubles on Moral Twin Earth: Moral Queerness Revisited», *Synthese*, 92 (1992), pp. 221-260.

las propiedades normativas o evaluativas sobrevienen a propiedades no normativas, no evaluativas. Varias posturas metaéticas se comprometen con la sobrevenida normativa, pero ofrecen diferentes explicaciones de su origen. De acuerdo con el naturalismo ético, la sobrevenida se da porque las propiedades normativas son definibles en términos de propiedades no normativas, naturalistas; es decir, las propiedades normativas resultan ser propiedades naturalistas. Los intuicionistas éticos, como G. E. Moore, verían la sobrevenida normativa como un hecho primitivo sintético *a priori* no susceptible de ulterior explicación; es algo que captamos directamente a través de nuestro sentido moral. R. M. Hare, un no-cognitista, intentaría explicarlo como una forma de una condición de consistencia que es esencial en el carácter regulativo del lenguaje de mandatos y prescripciones. Otros aún pueden intentar explicarla como procedente de la misma idea de evaluación normativa y defender que las propiedades evaluativas o normativas deben tener criterios descriptivos. Por consiguiente, está claro que la tesis de la sobrevenida normativa no sirve por sí misma para caracterizar una posición distintiva dentro de la metaética.

De manera parecida, es práctico tomar las diversas teorías mente-cuerpo como teorías que ofrecen acerca de la sobrevenida mente-cuerpo explicaciones que compiten entre sí. La explicación ofrecida por el fisicalismo de tipos es paralela a la explicación naturalista de la sobrevenida normativa: la sobrevenida mente-cuerpo se da porque lo mental es reducible físicamente, y las propiedades mentales, en última instancia, resultan ser propiedades físicas. El emergentismo, al igual que el intuicionismo ético, considera la sobrevenida mente-cuerpo como un hecho bruto no susceptible de explicación, algo que se debería aceptar, como insistió Samuel Alexander, con «piedad natural». Por el contrario, el epifenomenalismo recurre a la relación causal (al principio de que a la misma causa le sigue el mismo efecto) para explicar la sobrevenida, y en el funcionalismo, como veremos, la sobrevenida mente-cuerpo es una consecuencia de la opinión de que las propiedades mentales son propiedades funcionales que tienen como realizadores a propiedades físicas.

Debemos concluir, por tanto, que la sobrevenida mente-cuerpo no es en sí misma una *teoría explicativa*: simplemente establece un patrón de covariación de propiedades entre lo mental y lo físico, y señala la existencia de una relación de dependencia entre ambos dominios. Con todo, permanece completamente en silencio sobre la naturaleza de la relación de dependencia que podría explicar por qué lo mental sobreviene a lo físico. La sobrevenida no es una relación explicativa, metafísicamente profunda: es simplemente una relación fenomenológica acerca de patrones de covariación de propiedades. La sobrevenida mente-cuerpo, por lo tanto, *formula* el problema mente-cuerpo; no constituye una respuesta al mismo.

III

El dualismo cartesiano de sustancias describe el mundo como consistente en dos esferas independientes, la mental y la material, cada una con sus propias propiedades distintivas definitorias. Si bien hay interacciones causales a través de los dominios, las entidades de cada dominio, al ser «sustancias», son ontológicamente independientes entre sí, y es metafísicamente posible que un dominio exista en total ausencia del otro. Esta descripción de un mundo dicotomizado ha sido sustituida por el conocido modelo estratificado que ofrece una imagen del mundo como estratificado en diferentes «niveles», «órdenes» o «gradas», organizado en una estructura jerárquica. Se cree usualmente que el nivel inferior consiste en partículas elementales, o cualquier cosa que nuestra mejor física nos diga que son los elementos básicos de materia de los que se componen todas las cosas materiales¹². Según se sube en la escala, uno se va encontrando átomos, moléculas, células, organismos superiores, etc. La relación de orden que genera la estructura jerárquica es la relación mereológica: las entidades pertenecientes a un nivel dado, excepto aquellas del nivel más básico, se descomponen exhaustivamente, sin residuo, en entidades pertenecientes a los niveles inferiores. Las entidades en el nivel fundamental no tienen partes propias físicamente significativas.

¿Qué pasa entonces con las propiedades de estas entidades? Es parte de esta concepción estratificada el que en cada nivel se piense que haya propiedades, actividades y funciones que aparecen por primera vez en ese nivel (podemos llamarlas «las propiedades características» de ese nivel). Así, entre las propiedades características del nivel molecular están la conductividad eléctrica, la inflamabilidad, la densidad, la viscosidad y otras por el estilo; las actividades y funciones como el metabolismo y la reproducción están entre las propiedades características de los niveles celulares y biológicos superiores; la conciencia y demás propiedades mentales aparecen en el nivel de organismos superiores. Durante mucho tiempo a lo largo de este siglo, tal mundo estratificado ha conformado un trasfondo omnipresente, aunque sólo sea implícitamente, en el debate sobre el problema mente-cuerpo, la emergencia, el reduccionismo, el estatuto de las ciencias especiales y asuntos relacionados. Además ha ejercido una influencia generalizada en el modo en que formulamos los problemas filosóficos y debatimos sus soluciones. En ocasiones, el modelo estratificado se expresa en términos de conceptos y lenguajes en lugar de entidades del mundo y sus propiedades. Hablar de niveles de *descripciones*, niveles de *análisis*, niveles de *conceptos*, niveles de *explicaciones* y

12. Desde luego, el modelo a capas como tal no necesita postular un nivel básico; es consistente con una serie infinitamente descendente de niveles.

cosas semejantes es rampante en todas partes: ha impregnado completamente tanto la literatura científica primaria como los escritos filosóficos sobre la ciencia¹³.

Llegamos ahora a la cuestión crucial: ¿cómo se relacionan las propiedades características de un nivel dado con las propiedades de niveles adyacentes –en particular, con aquellas de los niveles inferiores–? ¿Cómo se relacionan las propiedades biológicas (vitales) con las propiedades psicoquímicas? ¿Cómo se relacionan la conciencia y la intencionalidad con las propiedades biológico/físicas? ¿Cómo se relacionan los fenómenos sociales, los fenómenos característicos de los grupos sociales, con los fenómenos que involucran individuos? Estarán de acuerdo en que éstas son algunas de las cuestiones centrales en filosofía de la ciencia, metafísica y filosofía de la mente. Las respuestas posibles a estas preguntas definen las opciones filosóficas clásicas en cada uno de los temas implicados. Entre las alternativas principales y mejor conocidas se incluyen el reduccionismo, el antirreduccionismo, el individualismo metodológico, el funcionalismo, el emergentismo, el neovitalismo y otras por el estilo. Se puede intentar dar una respuesta única e uniforme aplicable a todos los pares de niveles adyacentes, o se pueden tomar diferentes posturas según niveles diferentes. Se podría argumentar, por ejemplo, que las propiedades de cada nivel (que sea superior al nivel básico) son reducibles, en algún sentido bastante claro y sustancial, a propiedades del nivel inferior, o se podría restringir la pretensión reduccionista a ciertos niveles seleccionados (por ejemplo, propiedades biológicas respecto de propiedades psicoquímicas) y defender una postura antirreduccionista sobre las propiedades de otros niveles (por ejemplo, propiedades mentales). Además, no es necesario dar una respuesta uniforme respecto de todas las propiedades características de un nivel dado: respecto de las propiedades mentales, por ejemplo, es posible defender (y algunos han hecho precisamente eso) que las propiedades fenoménicas o sensoriales, o *qualia*, son irreducibles, mientras que sostienen que las propiedades intencionales, actitudes proposicionales incluidas, son reducibles (por ejemplo, funcional o biológicamente).

Volvamos ahora al enfoque reduccionista sobre la cuestión de las relaciones de propiedades entre niveles. Como ya he dicho, el reduccionismo, en particular el reduccionismo mente-cuerpo, sufrió deserciones masivas durante los 70 y los 80, con lo que apenas quedan, en parte alguna, reduccionistas a la vista en filosofía de la mente¹⁴. En general, creo que esto es verdad en todas las áreas de la filosofía;

13. En su obra sobre la visión David MARR distingue estupendamente bien tres niveles de análisis: el computacional, el algorítmico y el de implementación. Véase *Vision*, New York, Freeman Press, 1982. Los emergentistas, al inicio de este siglo, parecen haber sido los primeros en dar una formulación explícita del modelo a capas; véase por ejemplo C. LLOYD MORGAN, *Emergent Evolution*, London, Williams and Norgate, 1923. Para una formulación especialmente clara y útil, véase OPPENHEIM, Paul y Hilary PUTNAM, «Unity of Science as a Working Hypothesis», en Feigl MAXWELL y SCRIVEN (eds.), *Minnesota Studies in Philosophy of Science*, vol. II.

14. Andrew MELNYK escribe: «Por supuesto, parece haber una poco conocida ley que gobierna la conducta de los filósofos contemporáneos de que, en cualquier momento en que profesan su fe en cualquier forma de materialismo o fisicalismo, deben aclarar totalmente que, desde luego, no aprueban de

quizá queden aún reduccionismos o programas reduccionistas (y creo que efectivamente los hay), pero no conozco a nadie que se declare a sí mismo reduccionista sobre nada. Sin embargo, para empezar, ¿qué es reducción? Y ¿qué ha hecho del reduccionista una persona *non grata* en filosofía de la mente?

El concepto de reducción que ha servido como trasfondo compartido en la discusión del reduccionismo físico se derivó de un modelo de reducción elaborado por Ernest Nagel en los 50¹⁵. Nagel estaba principalmente interesado en la reducción interteórica, como relación entre dos teorías científicas. Su principal foco de interés estaba en la relación lógica entre la teoría que iba a ser reducida y la teoría que servía como base de reducción. De acuerdo con Nagel, la reducción es fundamentalmente un procedimiento de prueba consistente en la derivación lógico/matemática de las leyes de la teoría reducida a partir de las leyes de la teoría base, en conjunción con las «leyes puente» que conectan los predicados de ambas teorías. Nagel pensaba que estos enlaces interteóricos eran necesarios para asegurar las conexiones lógico/derivacionales entre ambas teorías, ya que las teorías pueden ser expresadas mediante vocabularios descriptivos completamente distintos. Normalmente, estas leyes puente son formalmente un tipo de bicondicionales (enunciados «si y sólo si»), que proporcionan para cada propiedad del dominio de la teoría reducida una propiedad nomológicamente coextensiva en la base de reducción. Así, en el caso de la reducción mente-cuerpo, el modelo de Nagel requiere que a cada propiedad mental corresponda una propiedad física nomológicamente coextensiva; es decir, tiene que ser válida una ley con la siguiente forma para cada propiedad mental *M*:

$$(BL) M \leftrightarrow P$$

donde *P* es alguna propiedad física.

Este requisito de la ley puente hizo del reduccionismo mente-cuerpo (de hecho de todos los reduccionismos) una diana fácil. Como señalé anteriormente, el argumento antirreduccionista más influyente, aquel que tuvo un papel decisivo en el establecimiento del consenso antirreduccionista, fue el argumento de la realización múltiple, que se basa en la observación de que, debido a su múltiple realizabilidad, las propiedades mentales no pueden tener coextensiones en el ámbito físico, de ahí que no se disponga de leyes puente mente-cuerpo para llevar a cabo la reducción nageliana. Posteriormente, este argumento se extendió para

modo alguno algo tan poco sofisticado, reaccionario y generalmente intolerable como el reduccionismo», en «Two Cheers for Reductionism: Or, the Dim Prospects for Non-Reductive Materialism», *Philosophy of Science*, 62 (1995), pp. 370-388. De acuerdo con Melnyk sólo quedan en escena dos reduccionistas; y dice: «La ley se da *ceteris paribus*; por ejemplo, no se aplica si tu nombre es Jaegwon Kim o Patricia Churchland».

15. Véase *The Structure of Science*, New York, Harcourt, Brace & World, 1961, capítulo 11. El modelo había sido desarrollado en artículos más tempranos de Nagel, publicados durante los 50.

defender una posición antirreduccionista general respecto de todas las ciencias especiales¹⁶. Esto ha hecho de las leyes puente el punto central de los debates sobre reducción y reduccionismo. Durante tres décadas las batallas sobre el reduccionismo se han librado en torno a la cuestión de si se dispone de leyes puente bicondicionales que conecten el dominio mental con el físico.

Pero éste es el campo de batalla equivocado para atacar el problema de la reducción. El modelo de Nagel es el modelo equivocado de reducción para las discusiones sobre la reducción mente-cuerpo, y las leyes puente son conexiones del tipo equivocado para inducir la reducción. Mi posición es que *las leyes puente no son ni necesarias ni suficientes para la reducción*. Creo que es fácil ver que la derivación mediante leyes puente no es suficiente para la reducción. Hay dos razones para ello. La primera tiene que ver con la importancia explicativa de una reducción: una reducción debe explicar cómo y por qué los fenómenos reducidos (los fenómenos «de nivel superior») surgen de los procesos en el nivel de la base de reducción (fenómenos «de nivel inferior»). Sin embargo, este requisito explicativo no se cumple cuando, como en el caso de la reducción de Nagel, las leyes puente se asumen como elementos primitivos e inexplicados de las derivaciones reductivas. Una ley puente del tipo (BL) sólo nos dice que la propiedad mental *M* (por ejemplo, dolor) co-ocurre, como una cuestión de necesidad nómica, con una propiedad física *P* (por ejemplo, activación de fibras C), pero la reducción de Nagel simplemente no se ocupa de la cuestión de por qué esto es así. ¿Por qué se experimenta dolor, en vez de picor o cosquillas, cuando se activan las fibras C? ¿Por qué no se experimenta dolor cuando se activan las fibras A-delta? ¿Por qué surge cualquier experiencia consciente cuando se activan estas fibras neuronales?

Cuando los emergentistas afirmaron que las propiedades de la consciencia son propiedades emergentes irreducibles, fue porque perdieron la esperanza de responder alguna vez a estas cuestiones explicativas. Ellos aceptaron tanto una ontología fisicalista fundamental como la sobreveniencia de las propiedades de nivel superior sobre las de nivel inferior; pero no se preocuparon de la múltiple realizabilidad de las primeras respecto de las segundas. La *disponibilidad* de las leyes bicondicionales de correlación era la menor de sus preocupaciones. Era la *inteligibilidad* de estas leyes lo que inquietaba a los emergentistas. Nos recomendaron que aceptáramos, con natural devoción, como hechos brutos el fenómeno de la emergencia, codificado en nuestras leyes puente. Hasta donde afectaba a los emergentistas, se nos invitaba a servirnos tanto como quisiéramos de la reducción nageliana de lo mental, pero nada más que un como ejercicio lógico: no se avanzaría ni un milímetro en nuestra comprensión de por qué, y cómo, hace su aparición lo mental cuando suceden ciertas configuraciones propicias de condiciones biológicas.

16. Véase J. A. FODOR, «Special Sciences (or The Disunity of Science as a Working Hypothesis)», *Synthese*, 28 (1974), pp. 97-115.

Alcanzar tal comprensión es exactamente la misma tarea que explicar los bicondicionales semejantes a (BL), es decir, las leyes puente mente-cuerpo.

Una segunda razón por la que no es suficiente la derivación nageliana *vía* leyes puente para la reducción es ontológica: esperamos que nuestras reducciones simplifiquen nuestro esquema de conceptos o esquema de identidades. No obstante, las leyes puente son consideradas normalmente como contingentes y empíricas, no como analíticas o *a priori*, lo cual significa que los conceptos *M* y *P* en (BL) siguen siendo distintos. Por tanto, la derivación de Nagel no contribuye en absoluto a la simplificación conceptual. Además, una ley puente expresa sólo la coextensividad nómica de las propiedades, no su identidad, lo cual significa que *M* y *P* siguen siendo propiedades distintas. Por tanto, la derivación de Nagel mediante las leyes puente tampoco contribuye a la simplificación ontológica. Si bien nos ofrece alguna simplificación de leyes (las leyes de la teoría reducida han sido absorbidas en la reductora), esto también podría ser en gran parte ilusorio, puesto que nos vemos forzados a inflar nuestra teoría base añadiendo nuevas leyes como leyes primitivas. La introducción de estas nuevas leyes puede representar una expansión significativa tanto de la ideología como de la ontología de la teoría base puesto que aquéllas traen consigo conceptos y propiedades ajenos a la teoría base original.

Llegados a este punto, la maniobra habitual dentro de la discusión filosófica en torno a la reducción de teorías consiste en considerar cómo y bajo qué condiciones correlaciones de la forma (BL) se pueden ampliar en identidades de la forma:

$$(I) M = P$$

Esta estrategia es apropiada siempre y cuando trabajemos dentro del paradigma nageliano. Creo, sin embargo, que la concepción nageliana de la reducción debería ser desechada. Si la reducción ha de seguir siendo un factor filosóficamente relevante dentro de nuestras reflexiones sobre el problema mente-cuerpo y sobre temas relacionados que conciernen a las relaciones de propiedades entre niveles, entonces nuestra concepción de la reducción necesita ser reorientada. Porque la pobreza filosófica de la reducción nageliana es fácil de ver. Como ya he señalado, no hay nada en el emergentismo que excluya una reducción *à la Nagel* de la psicología a la teoría física. Ni siquiera el dualismo de sustancias necesita excluir la reducción nageliana mente-cuerpo. Además, algunas formas de dualismo *implican efectivamente* la reducibilidad nageliana de la psicología a la teoría física: por ejemplo, la teoría del doble aspecto y la doctrina de la armonía preestablecida nos proporcionarían cada una de ellas un abundante número de leyes puente como para hacer posible una reducción nageliana mente-cuerpo (¡y reducciones en la dirección opuesta también!). Está claro, entonces, que cualquier doctrina reduccionista mente-cuerpo, expresada según la concepción nageliana de la reducción, no puede ser una tesis significativa sobre el estatuto de la mente. Si esto es así, la refutación del

reduccionismo mente-cuerpo en ese sentido de reducción no se puede considerar tampoco como un logro filosófico significativo.

IV

Esbozaré ahora un modelo de reducción que creo que es más adecuado tanto para la ciencia como para la filosofía. Si M y P en la ley puente (BL) son propiedades intrínsecas, la correlación entre M y P debe ser considerada como un hecho bruto sobre dos propiedades intrínsecas distintas, y ningún juego de prestidigitación filosófica puede convertirlo en una identidad. La única manera de ir más allá de tales correlaciones brutas es interpretar o reinterpretar el objetivo de la reducción, M , como una propiedad extrínseca/relacional. Veamos algunos ejemplos. Tómese el caso de la temperatura: para reducir la temperatura, debemos en primer lugar pensar en ella de manera relacional y caracterizarla en términos de su relación a otras propiedades. La temperatura es esa propiedad de un objeto, o sistema, cuya magnitud se incrementa cuando el objeto está en contacto con otro objeto de magnitud más alta que él; cuando es suficientemente alta, puede hacer que la madera o el carbón se quemen; cuando es extremadamente alta puede conducir al hierro al estado de fusión; cuando es suficientemente baja el agua se puede congelar; bien, se capta la idea. Lo que se hace es entender la temperatura como una propiedad caracterizada en términos de sus relaciones nomológico-causales con otras propiedades; es decir, se ofrece una caracterización extrínseca en cuanto «papel causal». Considérese otro ejemplo: la reducción del gen. Para empezar debemos analizar el concepto de gen en términos de su función causal: el gen es ese mecanismo en un organismo que es causalmente responsable de la transmisión de características heredables. La reducción de la temperatura se consigue cuando podemos identificar la propiedad que cumple la especificación causal: para los gases la propiedad resulta ser la energía cinética media de las moléculas; para los sólidos y plasmas, y en el vacío, resultan ser propiedades diferentes. La reducción del gen se lleva a cabo cuando identificamos el mecanismo que cumple el papel causal especificado: resulta ser la molécula de ADN, al menos, en los organismos terrestres.

Según esta concepción¹⁷, por tanto, la reducción de una propiedad M se lleva a cabo en dos pasos: (i) el paso conceptual de análisis de M en términos de sus

17. Las ideas aquí implicadas se retrotraen hasta David LEWIS, «An Argument for the Identity Theory», *Journal of Philosophy*, 63 (1966), pp. 17-25 y el argumento de David ARMSTRONG, a favor del materialismo de estado central en su *A Materialist Theory of Mind*, New York, Humanities Press, 1968. Véase igualmente Robert VAN GULICK, «Nonreductive Materialism and the Nature of Intertheoretical Constraint», en BECKERMANN, FLOHR y KIM (eds.), *Emergence or Reduction?* y Joseph LEVINE, «On Leaving What is Like», en W. Humphreys GLYN y Davies MATIN (eds.), *Consciousness*, Oxford, Blackwell, 1993. Son también importantes la discusión de David CHALMERS, de la «Explicación reductiva», en *The Conscious Mind*, New York, Oxford University Press, 1996, y los puntos de vista de Frank JACKSON, sobre el papel del análisis conceptual en metafísica, por ejemplo en «Armchair Metaphysics», en J. O'LEARY HAWTHORNE y M. MICHAEL (eds.), *Philosophy of Mind*, Dordrecht, KLUWER, 1993.

relaciones nomológico-causales con otras propiedades; y (ii) el paso empírico-teórico de identificar los «realizadores» de M , es decir, las propiedades, o mecanismos, en el dominio de la base de reducción que tienen las características nomológico-causales especificadas. Podemos esperar que el segundo paso involucre una teoría que explique precisamente cómo estos realizadores tienen estas propiedades nomológico-causales (una teoría de este tipo estará involucrada de manera casi segura en el proceso de identificación de los realizadores de las propiedades funcionales elegidas). El paso (i) consiste efectivamente en el proceso de «funcionalizar» la propiedad elegida, es decir, en el proceso de definirla en cuanto papel causal. En concreto, debemos pensar en la funcionalización en términos de propiedades de segundo orden: tener M es tener la propiedad de segundo orden de tener alguna propiedad, Q , que, a su vez, cumple la especificación C . Dado que esta última propiedad implica cuantificación sobre propiedades de primer orden (es decir, las propiedades ya dadas), se trata de una propiedad de segundo orden. Cuando la especificación C implica relaciones causales nómicas, podemos llamar a la propiedad de segundo orden propiedad «funcional»¹⁸.

En el paso (ii), por regla general, se espera tener realizadores múltiples. Por tanto, se debe afrontar la siguiente cuestión: ¿presenta también el fenómeno de la realizabilidad múltiple de la propiedad elegida dificultades para esta versión de la reducción? Responder a esta cuestión es un asunto algo complicado, que depende, hasta cierto punto, de una decisión sobre lo que nosotros queramos denominar «reducción». No obstante, se puede ofrecer un argumento persuasivo a favor de una respuesta negativa: no hay por qué tener miedo a la realización múltiple. Supongamos que M tiene dos realizadores, Q_1 y Q_2 . Para una cosa x , tener M es tener Q_1 o Q_2 . (Nótese que esto no introduce una propiedad disyuntiva: el «o» de tener Q_1 o Q_2 forma una oración disyuntiva, no un predicado disyuntivo con propiedades disyuntivas como valores semánticos). Es decir, cada instancia de M es una instancia de Q_1 o una instancia de Q_2 , y no hay ninguna instancia de M por encima de estas instancias de Q . Supongamos que M es instanciado en una ocasión en virtud de que su realizador Q_1 se instancia en esa ocasión. Esta instancia de M , entonces, es idéntica a esta instancia de Q_1 , y ambas tienen exactamente idénticos poderes causales: ningún poder causal nuevo puede mágicamente corresponder a la instancia de M que la instancia de Q_1 no posea. Los realizadores de M hacen todo el trabajo que hace M , y M misma no representa una adición neta a la ontología o la estructura causal del mundo.

18. La noción de propiedad de segundo orden en este sentido se debe a Hilary PUTNAM, «On Properties», en N. RESCHER *et al.* (eds.), *Essays in Honor of Carl G. Hempel*, Dordrecht, Reidel, 1969. Es interesante resaltar que, aunque el inventor del funcionalismo también introdujo el concepto de una propiedad de segundo orden, que está hecha a medida para una perspicua explicación de «realización», ningún funcionalista, en lo que yo sé, se aprovechó de ella hasta que lo hizo Ned BLOCK en su «Can the Mind Change the World?», en George BOLOS (ed.), *Meaning and Method*, Cambridge, Cambridge University Press, 1990.

Sin embargo, alguien podría objetar: «¿Pero qué ocurre con *M misma*? *M* no es idéntica ni a Q_1 ni a Q_2 , y por lo tanto se debe considerar como una propiedad distinta a todas y cada una de las propiedades del dominio de reducción». Ésta es el mismo reto ontológico que planteamos a la reducción nageliana: ¿dónde está el beneficio ontológico en lo que se refiere a *M*? ¿Se requiere que *M* siga aún mero-deando en nuestra ontología? Creo que podemos tratar *M* de una de estas dos maneras. Una manera simple es identificar *M* como la propiedad disyuntiva, $Q_1 \vee Q_2$. Si las Q s son diversos realizadores de *M*, su diversidad debe significar algo, y la única cosa que podría significar es la diversidad nomológico-causal. Si son idénticas causal y nomológicamente o bastante similares, no habría razón para considerarlas como realizadores distintos. Generalmente se acepta que las clases en ciencia son individuadas primariamente en base a sus poderes causales. Así *M* como una disyunción de propiedades causalmente diversas será una clase causalmente heterogénea, y tendrá sólo utilidad limitada como clase científica¹⁹.

El segundo modo de tratar *M* es considerarla sólo como un *concepto*, no como una propiedad. Mediante la formación de una expresión de segundo orden de la forma «tener una u otra propiedad, Q , tal que $C(Q)$ », no podemos introducir literalmente una nueva entidad en nuestra ontología. Todo lo que hacemos es introducir una manera de distinguir ciertas propiedades de primer orden mediante la especificación de una condición que necesitan cumplir; podríamos decir que una expresión de segundo orden de esta forma se refiere indiscriminadamente a miembros de una clase de propiedades de primer orden, es decir, a aquellas que satisfacen la condición especificada. Mediante meras operaciones lingüísticas como la cuantificación no podemos ni expandir ni contraer nuestra ontología; lo que expandimos es nuestro repertorio lingüístico.

Ésta, creo, es una respuesta suficiente a la pregunta ontológica. Muestra cómo una reducción funcional nos proporciona una ontología simplificada. Pero, ¿cómo cumple el modelo funcional de reducción los requisitos explicativos de la reducción? ¿De qué modo una reducción funcional es una reducción explicativa? Creo que esta pregunta tiene una respuesta satisfactoria. ¿Por qué *M* es instanciada en los sistemas del tipo *S* siempre que Q_i es instanciada por estos sistemas? Porque tener *M* sólo es tener alguna propiedad que cumple la especificación causal C , y Q_i es una propiedad que realiza *M* —es decir, cumple la especificación C — en sistemas del tipo *S*. ¿Por qué este sistema particular instancia *M* en esta ocasión? Porque está instanciando Q_i , uno de los realizadores de *M*. ¿Por qué esta instancia *M* causa un efecto de la clase *E*? Porque es, de hecho, una instancia Q_i , donde Q_i es un realizador de *M*, y las instancias de Q_i tienen efectos de la clase *E*. Dado que los poderes causales de las instancias de *M* se identifican con los de sus realizadores, todas las preguntas sobre las relaciones

19. Argumento que tales propiedades no son proyectables inductivamente en «Multiple Realization and the Metaphysics of Reduction», en *Supervenience and Mind*, Cambridge, Cambridge University Press, 1993.

causales que involucran instancias de M admiten respuesta en el nivel de los realizadores de M . ¿Qué más podemos exigir a una reducción explicativa de M ?

Cuando se introdujo el funcionalismo como una alternativa al clásico fisicalismo de tipos, es decir, a la teoría del estado cerebral, se pensó, y aún se piensa generalmente, que era una forma de antirreduccionismo, de hecho, la versión principal del antirreduccionismo sobre lo mental. Por lo que estoy abogando es precisamente por lo contrario: *la funcionalizabilidad de las propiedades mentales es necesaria y suficiente para la reducción (a la espera de un descubrimiento científico exitoso de sus realizadores para dominios que nos interesen)*. Esto no es simplemente una redefinición del término «reducción»; espero haberles persuadido de que el modelo funcional nos conduce hacia el modo correcto de pensar sobre la reducción. En este modelo de reducción, entonces, las propiedades emergentes son fácilmente caracterizadas: una propiedad M es emergente relativa a un dominio dado D de propiedades sólo en el caso de que M no sea funcionalizable en términos de propiedades de D .

V

Para evaluar dónde nos encontramos ahora con el problema mente-cuerpo, debemos, por tanto, saber dónde nos situamos en relación con el planteamiento funcionalista hacia lo mental. Se suele distinguir entre dos grandes categorías de fenómenos mentales, lo intencional y lo fenoménico, sin excluir aquellos fenómenos que comparten aspectos de ambos (por ejemplo, las emociones). La intencionalidad es especialmente evidente en las actitudes proposicionales como la creencia, el deseo y la intención. La viabilidad de una versión funcionalista de la intencionalidad ha despertado mucho escepticismo; concretamente, Hilary Putnam, el padre del funcionalismo, ha lanzado recientemente sostenidos ataques contra los modelos causal-funcionalistas del contenido y la referencia, y John Searle también se ha opuesto con vigor a la funcionalización de la intencionalidad²⁰. Sin embargo, estos argumentos siguen sin convencerme; no veo obstáculos insuperables para una versión causal-funcional de la intencionalidad. Permítaseme sólo mencionar aquí que me parece inconcebible que exista un mundo posible que sea un duplicado físico exacto de este mundo, pero que carezca completamente de intencionalidad²¹. Un mundo así tiene que ser idéntico al nuestro en todos los aspectos psicológico-intencionales²².

La dificultad llega con los *qualia*. Pues, a diferencia del caso de los fenómenos intencionales, parece que podemos concebir un duplicado físico de este

20. Véase Hilary PUTNAM, *Representation and Reality*, Cambridge, Ma., MIT Press, 1988; Jhon SEARLE, *The Rediscovery of Mind*.

21. Creo que otros (quizá Shoemaker y Block) han hecho una observación similar.

22. Una posición como ésta la defiende explícitamente David CHALMERS en *The Conscious Mind*.

mundo en el que los *qualia* se distribuyan de manera diferente o estén completamente ausentes (un «mundo de zombis» como algunos lo llaman). Para ir al grano sin ceremonias, me parece que las cualidades sentidas, fenoménicas, de las experiencias, o *qualia*, son propiedades intrínsecas si algo lo es. Efectivamente, nos referimos con frecuencia a ellas usando descripciones causales extrínsecas: por ejemplo, «el color del jade», «el olor del amoníaco», «el sabor del aguacate», etc. Sin embargo, esto es completamente consistente con la afirmación de que lo que todas estas descripciones recogen son cualidades intrínsecas, no algo extrínseco ni relacional. (Se puede argumentar que es a causa de que son intrínsecas y subjetivas por lo que necesitamos acudir a descripciones relacionales para una referencia intersubjetiva). Compárese con nuestra práctica de adscribir propiedades físicas intrínsecas a objetos materiales mediante el uso de descripciones relacionales; por ejemplo, «dos kilos», «32 grados Fahrenheit», etc. Decir que un objeto tiene una masa de 2 kilos es decir que, en una balanza de brazos iguales, equilibrará dos objetos, cada uno de los cuales equilibraría el Kilo Prototipo (un objeto almacenado en algún lugar de Francia). Ése es el significado lingüístico, el «concepto» si se prefiere, de «2 kilos»; sin embargo, la propiedad que designa, tener una masa de dos kilos, es una propiedad intrínseca de los cuerpos materiales.

Si las propiedades cualitativas de la consciencia son intrínsecas, resistirán la funcionalización y, por ello, la reducción. Mis dudas sobre las versiones funcionalistas de los *qualia* se basan por lo general en los bien conocidos, aunque no incontestados, argumentos a partir de las inversiones de *qualia* y de las habituales consideraciones epistémicas. En cualquier caso me parece que si el emergentismo es correcto sobre algo, es más probable que lo sea sobre los *qualia* que sobre cualquier otra cosa²³.

Esto es lo que hace que la postura que se tome en el problema de los *qualia* represente un punto decisivo de elección con respecto al problema mente-cuerpo. Permítaseme acabar haciendo notar cómo la cuestión de la reducibilidad se relaciona con otro problema central en la metafísica de la mente, el llamado problema de la causación mental. Si la propiedad mental *M* es funcionalmente reducible en nuestro sentido, hay, entonces, una fácil respuesta a la cuestión de cómo *M* puede tener poderes causales en el dominio físico. Como ya hemos señalado, los poderes causales de cualquier instancia dada de *M* son idénticos a los poderes causales del realizador particular físico de *M* en esa ocasión, y no hay ninguna adición neta de poderes causales más allá de los de las propiedades físicas. Pero si *M* no es funcionalmente reducible, es difícil, de hecho no es posible, ver cómo *M* o las instancias de *M* pueden ejercer poderes causales en el dominio físico si asumimos, como creo que deberíamos hacer, que el dominio físico está

23. Esta posición sobre los *qualia* y el reduccionismo tiene una semejanza cercana a las posiciones defendidas por numerosos filósofos, en concreto Joseph Levine, Frank Jackson, David Chalmers y quizá Ned Block.

causalmente cerrado. Así, el precio que podemos tener que pagar (creo que es el que debemos pagar) por la irreducibilidad de los *qualia* es el de sus poderes causales: si son irreducibles, se ven amenazados de impotencia causal, al menos, en el dominio físico²⁴.

De esta manera, los problemas centrales de la filosofía de la mente, el problema de la conciencia y el de la causación mental, se reúnen en el mismo foro. La única manera visible de explicar la conciencia físicamente –es decir, de encontrar para ella un lugar en el mundo físico– es funcionalizarla en el ámbito físico. Si se pudiera hacer eso, se podría resolver también el problema de sus poderes causales. Si, como parece probable, no se puede hacer, la conciencia está amenazada por el epifenomenalismo. Parece, por tanto, que podemos preservar la conciencia, o cualquier otro aspecto de nuestra mentalidad, como algo distintivo y autónomo únicamente si estamos dispuestos a aceptar su impotencia causal. En pocas palabras, los dos problemas se hacen uno a otro irresolubles²⁵.

Creo que cincuenta años de debate han demostrado que el núcleo principal del problema mente-cuerpo está constituido por dos grandes y profundos rompecabezas, el de la conciencia y el de la causación mental. Y estos dos rompecabezas resultan estar íntimamente entrelazados: la clave para ambos es la cuestión de si las propiedades fenoménicas de la conciencia se pueden funcionalizar o no. Creo que es aquí donde nos encontramos ahora en el problema mente-cuerpo, medio siglo después de su reintroducción en la filosofía a cargo de Ryle, Smart, Feigl y otros.

24. Creo que la irreducibilidad conduce a la impotencia causal *tout court*, pero un argumento detallado debe esperar a otra ocasión.

25. Este modo de plantearlo me fue sugerido por David Chalmers en conversación.