

Epistemología, filosofía de la mente y bioética

En busca de la mente cerebral.

Del alma al *software* (2)*

Felipe de Brigard**

A pesar de sus valiosos cambios teóricos y de la evidente ventaja de su metodología, el conductismo fue prontamente criticado, y hacia la tercera década del siglo XX su popularidad empezó a decrecer. Una de las más célebres refutaciones al conductismo lo acusa de circularidad, producto quizás de la inexactitud del concepto *disposición conductual* (1). La idea, en breve, es que los análisis en términos disposicionales conductuales —expresados en oraciones hipotéticas—, no eliminan la atribución de conceptos mentalistas: más bien, los presuponen. Supongamos, por ejemplo, que queremos hacer un análisis de este estilo de la aserción psicológica “Pedro tiene dolor de cabeza”. De acuerdo con el conductismo lógico, esta proposición podría ser traducida a una lista de oraciones hipotéticas comportamentales, entre las cuales podrían aparecer:

- a) Si Pedro tuviera a la mano una aspirina, entonces la tomaría.
- b) Si le preguntaran a Pedro: “¿qué le ocurre?”: él diría “me duele la cabeza”.
- c) Si a Pedro le ofrecieran un masaje para aliviar el dolor, él lo aceptaría.

Esta lista podría ser incluso infinita; no obstante —sostendría Ryle—, en principio esto no sería un problema. Él era consciente de que, de hecho, “hay muchas disposiciones cuyas actualizaciones pueden tomar una amplia y quizás ilimitada variedad de formas” (2); a pesar de ello, creía que en cuanto existieran algunos conceptos disposicionales verificables, el problema se resolvería. Pero el argumento de Putnam (1) apunta, más bien, a que para que la reducción resulte válida, a), b) y c) deben presuponer una serie de oraciones que involucran conceptos mentales. Sólo para el caso de justificar a) debe suponerse, entre otras cosas, que:

.....
* En la primera parte del artículo, publicado en el número 2 de este volumen, comenzamos revisando con cierto detalle los argumentos originales (cartesianos) a favor del dualismo, y terminamos analizando la línea argumentativa seguida por los partidarios del conductismo en su forma clásica. El presente escrito retoma las ideas del anterior y empieza con las más famosas críticas que, en su momento, se le formularon a la postura conductista. Los subtítulos y el número que acompaña a las tesis de cada una de las posturas, mantienen la coherencia con el ordenamiento iniciado en la primera parte.

** Pontificia Universidad Javeriana, Facultad de Filosofía.

- a₁) Pedro *desea* acabar con su dolor de cabeza.
- a₂) Pedro *cree* que la aspirina puede quitarle el dolor de cabeza.
- a₃) Pedro *cree* que las pastillas que tiene en la mano son aspirina y no otra cosa.

Algo similar ocurriría con b) y con c) y, eventualmente, con el resto de la posible lista de oraciones hipotéticas del análisis. Ahora bien, la eliminación del término en cursiva (el término intencional) en a₁, a₂ y a₃ tampoco parece factible, pues es apenas esperable que en el análisis de cada una de ellas vuelva a presentarse la misma situación. Resulta, por lo tanto, que el conductismo no elimina el lenguaje intencional.

Por otro lado, Putnam (3) esgrimió contra el conductismo otra importante crítica que, por así decirlo, apelaba al sentido común. Él creía que la intención principal del conductismo era enfrentarse a la concepción, tanto dualista como materialista, de que aunque “el significado de [un término mental como] “dolor” pudiese ser *explicado* por referencia al comportamiento observable, lo que queremos decir con “dolor” no es la presencia de un grupo de respuestas, sino más bien la presencia de un evento o condición que normalmente causa esas respuestas” (3). Pero consideraba que la eliminación absoluta del nexo causal entre la mente y la

conducta no podría ser posible. Para mostrarlo ideó un experimento mental que consistía, en pocas palabras, en lo siguiente: imagina, por un lado, un mundo poblado por “superactores” que pudieran imitar perfectamente la conducta de alguien que siente dolor, aunque no lo sintieran. De forma similar, concibe otro mundo habitado por “superespartanos” a quienes les enseñan, desde que nacen, a resistir cualquier tipo de dolor y a comportarse como si no lo sintieran. Estos superespartanos permanecen tan impresionantemente impasibles frente a los estímulos dolorosos que cualquiera creería que no sienten nada en lo absoluto. Su conclusión es que en cualquiera de estos mundos el conductismo resultaría inútil para explicar la naturaleza del dolor, pues le adscribiría dolor a los farisantes superactores, al tiempo que se lo negaría a los estoicos superespartanos. La razón, aduce, es que el dolor no puede ser interpretado meramente como una conducta que expresa dolor, sino como aquello que *causa* esa conducta. Pero tal causa es inapelable desde el conductismo.

Finalmente, es importante referirse también a una tercera crítica, esta vez centrada en los términos conductistas acuñados por Skinner, expuesta principalmente en la reseña de Chomsky a *Verbal Behavior* (4). Chomsky argüía, en contra del psicólogo, que mientras conceptos

como estímulo o respuesta permanecieran dentro de contextos experimentales tan específicos como el laboratorio, su uso podría resultar adecuado, mas a la hora de intentar aplicarlos en la vida diaria, terminan por ser definitivamente inapropiados. De acuerdo con Skinner, un evento ambiental sólo puede ser identificado como el estímulo que elicitaba una respuesta, en caso de que pueda establecerse una relación legaliforme que los conecte. Pero esta relación —sugiere Chomsky— no siempre es tan sencilla de establecer. Supongamos el caso de que nos mostrasen una pintura de una escuela holandesa. Una típica respuesta, pronunciada bajo el control del adecuado condicionamiento hacia un estímulo, sería: “holandés”. Pero supóngase —continúa Chomsky— que en lugar de decir *holandés* hubiéramos dicho *Cuadra con el papel de colgadura, Yo pensé que a usted le gustaba el arte abstracto, Nunca antes lo había visto, Está torcido, Cuelga demasiado bajo, Bello, Horrible, ¿Recuerdas nuestro campamento el último verano?*, o cualquier otra cosa que pudiese haber venido a nuestra mente cuando mirábamos la pintura... (4).

De acuerdo con el conductismo, Skinner tendría que demostrar que

cada una de estas respuestas está efectivamente condicionada por una u otra propiedad física distinta en la pintura, propiedad ésta que, dadas ciertas variables controladas, funciona como estímulo elicitor. Pero esta búsqueda resultaría horrorosamente tediosa y hasta artificial, a tal punto que el término *estímulo* perdería toda su objetividad (4). ¿Qué ocurriría —continuando con el ejemplo— si una misma respuesta, v.gr. “bello”, fuera pronunciada con la misma fuerza y la misma latencia en presencia de otro cuadro o de una pieza musical? ¿Habríamos de encontrar una propiedad física compartida por los tres objetos? Esto nos lleva a poner en duda la eficacia del concepto de estímulo, pues nunca queda claro, en realidad, cuál es el evento físico frente al cual el organismo reacciona que pueda ser catalogado como tal. De ahí que Chomsky afirme que el modo en que Skinner usa estos términos no parecería ser más que una “paráfrasis del vocabulario que comúnmente usamos para describir comportamientos, y que no tienen conexiones particulares con sus homónimas usadas en el laboratorio” (4). Tarde o temprano, termina concluyendo, “los estímulos dejarán de ser parte del ambiente y volverán al interior del organismo” (4)¹.

¹ Estoy convencido de que el argumento de Chomsky en contra del conductismo metodológico encontraría suficientes respuestas dentro de la obra de Skinner, pero considero innecesario hablar de ellas aquí.

Fisicalismo

Además de las anteriores, muchas han sido las críticas que durante los últimos cincuenta años se le formularon al conductismo. Todas ellas deben dejar en claro que no sólo no es fácil eliminar los términos intencionales de la psicología a través del recurso conductual, sino que la apelación a estados mentales para explicar nuestras conductas no es del todo una estrategia desdeñable. Y fue —entre otras cosas— en virtud de las críticas al conductismo como surgió la idea de que las explicaciones en psicología debían incluir la referencia a procesos internos que causaran los comportamientos, pero evitando en todo caso caer en algún tipo de dualismo. Esto dio paso a un segundo tipo de enfoque psicológico materialista, desarrollado principalmente por U.T. Place, H. Feigl y J.J.C. Smart, que comúnmente se conoce como *teoría de identidad de tipos* —se le llama “de tipos” (*type*) para distinguirla de la “teoría de identidad de instancias” (*token*)—. Aunque en un principio los partidarios de esta nueva doctrina parecían conformes con algunas de las explicaciones disposicionales del conductismo, creían que esta aproximación dejaba un “residuo de conceptos intratables agrupados alrededor de las nociones de conciencia, experiencia, sensación e imaginación mental, en cuyo caso algún tipo de historia sobre procesos internos es inevitable”

(5). Pero propusieron que tales procesos ocurrían en el cerebro y no, como podría sostenerlo un dualista, en algún tipo de sustancia espiritual.

Las razones que alimentaron esa concepción podrían dividirse, según veo, en cuatro principales: en primer lugar, el hecho de que es un enfoque inscrito dentro de un marco teórico claramente positivista —quizás, mucho más que el propio conductismo—, donde el lenguaje de la física sigue siendo el ideal de lenguaje científico. En segundo lugar, el deseo filosófico de aplicar la navaja de Occam en contra de cualquier teoría que postulara entidades innecesarias en sus explicaciones. Smart, por ejemplo, sostenía como un hecho que la ciencia nos está ofreciendo cada vez más explicaciones mecanicistas, y la idea de “que todo pueda ser explicable en términos de la física, excepto la ocurrencia de sensaciones [...] parece francamente increíble” (6). De ahí que crea que no hay buenos motivos para permitir un tipo de sustancia espiritual que explique un campo de eventos tan reducido como puede serlo el de la conciencia, las sensaciones y, en general, el de las experiencias que llamamos mentales. En tercer lugar, porque se consideraba que tampoco había buenas razones para esperar que todas las explicaciones psicológicas pudieran ser formuladas únicamente en términos conductistas. Al igual que Place,

Smart pensaba que muchas explicaciones en términos exclusivamente de conductas o de disposiciones por actuar, quedarían incompletas. Dice, por ejemplo, en una crítica abiertamente dirigida a Wittgenstein (7, §244), quien afirmaba que la expresión verbal del dolor realmente no *describía* nada, sino que *reemplazaba* un comportamiento doloroso, v.gr. un grito: “Me parece que [...] cuando alguien dice ‘tengo un dolor’, está haciendo algo más que ‘reemplazar un comportamiento doloroso’, y ese ‘algo más’ no es sólo decir que está en una pena” (6) (cursiva mía). Y, finalmente, por la unión del supuesto de que nuestra constitución es enteramente física, y las pruebas de que todos los procesos (conocidos) que tienen que ver con el control del comportamiento son cerebrales (8). Consecuentemente, dado que esos “procesos internos” de los que carecía el conductismo debían ocurrir en algún sustrato físico, y el mejor candidato para serlo era el cerebro, formularon su famosa tesis de identidad: los estados y procesos mentales son idénticos a estados y procesos del cerebro y del sistema nervioso central (tesis de la teoría de identidad de tipos).

Esta perspectiva motivó así un nuevo tipo de reduccionismo, en el que los términos mentales serían primero reducidos al lenguaje de la neurofisiología y, de ahí, al de la física. Es importante hacer dos aclaraciones respecto a la naturaleza de

esta tesis de identidad. Por un lado, ella no sostiene que deba darse una identidad de *sentido* entre los términos mentales y cerebrales, aunque sí una identidad de *referencia*, siguiendo la diferencia establecida por Frege (9) entre *Sinn* y *Bedeutung*. Argüían que su tesis no debía entenderse como si, por ejemplo, dolor connotara lo mismo que proceso cerebral de la clase *x* (donde *x* es reemplazada por una descripción de una cierta clase de proceso cerebral); sino más bien, como un reporte de un proceso que ocurre en el cerebro (6). Esto hace que la teoría no pueda ser rebatible por el hecho de que podamos predicar cosas de los procesos mentales que no podemos predicar de los cerebrales. Una crítica así podría surgir de la consideración de que si la teoría propone una identidad debe cumplir con la Ley de Leibniz, la cual establece que dos cosas son (numéricamente) idénticas si y sólo si cualquier propiedad que pertenezca a la primera pertenece también a la segunda. Pero, como afirma Lewis:

La teoría de identidad no implica que cualquier cosa que sea verdadera de las experiencias [mentales] como tales sea igualmente verdadera de los estados neuronales como tales, y viceversa. [...] No debemos identificar una experiencia en sí con el atributo que se predica de [sic. la experiencia de] alguien al decir que está teniendo esa experiencia (Lewis, 1984, 8).

De este modo, resulta válido que uno pueda decir, v.gr., “veo una manzana roja” o “tengo un dolor intenso” sin que tenga que seguirse que el estado neural idéntico al proceso que tal reporte describe sea en sí mismo rojo o intenso (10). Similarmente, puede afirmarse que un cierto estado cerebral, idéntico al proceso descrito en términos de recuerdos o emociones, ocupe un espacio determinado, aunque no pueda decirse lo mismo de la experiencia mental a la que éste refiere.

Esta distinción, además, libera a la teoría de la crítica del “campesino iletrado”. Dicha objeción, a grandes rasgos, sostiene que los procesos mentales no pueden ser cerebrales, porque un campesino iletrado puede no saber nada de neurofisiología y aún así hablar de sus sensaciones o emociones. La crítica parece desprenderse de algún partidario de la teoría del significado de *Fido-Fido*: si el significado de una expresión fuera simplemente lo que ella expresa, “se seguiría del hecho de que “sensación” y “proceso cerebral” tienen diferentes significados, que ellos no pueden nombrar una y la misma cosa” (6). Sucede con este caso —siguiendo el ejemplo de Place y Smart— lo que le ocurre a alguien que habla de los relámpagos sin saber que son, en realidad, descargas eléctricas. Que una persona no sepa que un relámpago *es* una descarga eléctrica, no significa que no *pueda serlo*. El

hecho de que dos expresiones tengan independencia lógica —dice Place (5)— no significa que deban tener independencia ontológica, pues, en este caso, es una verdad empírica que los relámpagos son (idénticos a) descargas eléctricas, y la ciencia así los explica: “Decimos que lo que un relámpago realmente es, lo que su naturaleza es como nos lo revela la ciencia, es una descarga eléctrica” (6). Y esto esclarece, igualmente, por qué dos expresiones que refieren a una misma realidad pueden requerir operaciones diferentes para ser verificadas: mientras la mera observación nos es suficiente para verificar la ocurrencia de un relámpago, necesitamos instrumentos especiales para descubrir que aquello que vimos es una descarga eléctrica. Análogamente, los métodos de la neurociencia nos demostrarán que las observaciones introspectivas reportadas por un sujeto son en realidad procesos que ocurren en el cerebro, por lo cual podrán ser explicadas en sus términos (5). Pero alguien podría, de nuevo, objetar tal afirmación asegurando que así como es posible que un relámpago *sea* una descarga eléctrica, también es posible que *pueda no serlo*. La respuesta es simple: claramente, ésta es una posibilidad lógica, pero resulta ser una imposibilidad empírica y, al igual que ocurre con la identidad relámpago-descarga eléctrica, la identidad mente-cerebro es *contingente*, i.e. depende de cómo sea la realidad. Y ésta,

precisamente, es la segunda característica de la tesis de la teoría de identidad a la que quería referirme.

La identidad entre los procesos-estados mentales y los cerebrales debe distinguirse de la identidad propuesta por los conductistas entre los procesos-estados y las disposiciones conductuales, pues mientras los segundos creían que era un hecho conceptual, una verdad analítica o a priori, que los enunciados mentales podían reducirse a enunciados sobre disposiciones conductuales, los primeros sostenían que su tesis de identidad era una verdad empírica *a posteriori*, que no se desprendía de la lógica misma de los términos. Suponían, por ello, que su tesis en realidad era una hipótesis contingente, verificable empíricamente y, por lo tanto, no susceptible de refutación lógica alguna (aunque, como veremos, es la lógica modal la que se encargará de darle la estocada final a esta teoría). Esto quiere decir que no habría nada autocontradictorio en decir “ x tiene un dolor, pero no ocurre nada en su cerebro” (5). De ahí que la tesis cartesiana de que los estados mentales son estados de una sustancia inmaterial, *podría haber sido verdadera*, ya que en sí misma no encierra ninguna contradicción; ocurre, simplemente, que es un hecho empírico que no lo sea. Similarmente, podría pensarse en la objeción de que podemos imaginarnos a nosotros mismos como teniendo

una configuración física distinta y, aun así, seguir experimentado los mismos deseos y dolores; sucede, de nuevo, que éste no es el caso.

Pero, ¿cómo debe entenderse esta “identidad contingente”? La explicación que nos ofrecen los partidarios de la teoría de identidad de tipos corre de la siguiente forma: cuando se dice que “un proceso mental *es* un proceso cerebral”, el “es” de la cópula puede entenderse, bien como conectando un sujeto y un predicado en una verdad necesaria, o bien como conectándolos en una contingente. En el primer caso hablamos de un “es” de *definición*, mientras que en el segundo nos referiremos a un “es” de *composición* (5). Así, en la afirmación “un cuadrado *es* un rectángulo equilátero”, la cópula indica una relación definicional y, por lo tanto, una identidad necesaria, i.e. se sigue de ser un cuadrado el ser un rectángulo equilátero, y no es posible que ningún rectángulo equilátero no sea a su vez un cuadrado, y viceversa. Sin embargo, en una oración como “mi mesa de noche *es* un viejo baúl”, la composición entre sujeto y predicado, expresada por el “es” de la cópula, enuncia una verdad contingente que debe ser verificada por observación. Pues no es una verdad necesaria que todas las mesas de noche sean viejos baúles, pero sí una verdad contingente el hecho de que la mía lo sea. Vivimos en un mundo donde es verdad siempre que los cuadrados son rectán-

gulos equiláteros, pero no en el que todas las mesas de noche sean baúles viejos; si viviéramos en un mundo así, “los dos conceptos no tendrían un estatuto lógico independiente” (5) y, consecuentemente, dondequiera que se aplicara un lado de la identidad, podría aplicarse el otro. Es, pues, en este segundo sentido —en el que la cópula es de composición— como deben entenderse las proposiciones de identidad mente-cerebro.

Si es cierta, la teoría de identidad de tipos sería, sin duda, una buena solución para los problemas suscitados por las doctrinas que la precedieron. Nos facultaría, por ejemplo, para evadir los contraargumentos del conductismo, pues si los estados internos son *eficazmente causales*, sería posible, por un lado, correlacionar dos estados cerebrales como *causando* un mismo tipo de comportamiento (cosa que, como vimos, los términos de Skinner no permitían) y, por el otro, “llegar a hacer inteligible que [un] estado pueda ocurrir algunas veces pese a la disimulación de sus manifestaciones definitorias” (10), lo cual era, justamente, el centro de la crítica de Putnam. Pero, a decir verdad, la teoría de identidad de tipos va más allá de la mera aseveración de que los procesos mentales son esencialmente físicos. Aceptarla nos lleva a apoyar la idea de que “la neurociencia descubrirá una taxonomía de los estados neurales que se ponga en correspondencia uno-a-uno con la

taxonomía que a partir del sentido común hacemos de los estados mentales” (8). Esto, aunque consistiría —como se vio— en un logro científico a posteriori, está amparado por el supuesto de que la neurociencia y, posteriormente, la física, constituyen marcos conceptuales que funcionan mejor para explicar y predecir fenómenos que los que podrían ser predichos en una psicología de otro estilo. De este modo, en caso de que tanto los principios como las proposiciones mismas de la psicología pudieran subsumirse bajo, y aplicarse a los mismos casos que, las explicaciones físicas, tendríamos muy buenas razones para creer que los términos de esta segunda ciencia describen mejor la realidad de nuestra vida mental (8).

A pesar de la evidente ventaja científica que para la psicología supondría la verdad de la tesis de identidad, ésta se vio prontamente invadida de críticas que terminaron por desmentirla. Quizás una de las más importantes aparece en boca de Saúl Kripke (11), quien estaba particularmente interesado en la lógica de los enunciados de identidad. Su argumento está encaminado a negar la posibilidad de identidades contingentes: precisamente, la clase de identidad a la que, se supone, pertenece la tesis reduccionista que acabamos de estudiar. Los partidarios de esta tesis suponían que la afirmación “un dolor es un estado cerebral” era contingente en el mis-

mo sentido en que podría serlo “el calor es el movimiento de las moléculas” (11). Ellos sostendrían la contingencia de la segunda identidad sobre la base de que es posible imaginar que hubiese calor aun cuando no hubiera un cierto movimiento de moléculas, sino, por ejemplo, cuando una sustancia calórica empapase a los objetos. De modo similar, presumen contingencia en la tesis de identidad de tipos por el hecho de que uno puede suponer que 1) exista un estado cerebral sin que haya dolor en lo absoluto y 2) que exista una criatura que no tenga cerebro, pero sí dolor. En otros términos, la idea de contingencia en estas identidades aparece porque “es posible imaginar circunstancias definidas en las que esta relación habría podido ser falsa” (11); empero, se acepta que tal hecho no dañaría en lo absoluto la posibilidad de la identificación entre el calor y las moléculas, o entre un estado mental como el dolor y un cierto estado cerebral.

Sin embargo, Kripke rebate la posibilidad de las identidades contingentes a partir de un profundo análisis del concepto de identidad desde la perspectiva de la lógica modal, cuyo argumento central intentaré reconstruir brevemente aquí (11). La ley de sustitutividad de idénticos sostiene que para cualquier par de objetos x e y , si x es idéntico a y , entonces si x tiene una propiedad F , y también la tiene: (1). $(x).(y) [(x = y) \rightarrow (Fx \rightarrow Fy)]$. Por

otra parte, se asume también que cualquier objeto es necesariamente idéntico a sí mismo: (2). $(x) \diamond (x = x)$. De ahí que pueda ser considerada como una sustitución de (1) la fórmula: (3). $(x).(y) (x = y) \rightarrow [\diamond (x = x) \rightarrow \diamond (x = y)]$. De forma que, eliminando la cláusula $\diamond (x = x)$ del condicional, que por (2) se acepta como verdadera, llegamos a la conclusión: (4). $(x).(y) [(x = y) \rightarrow \diamond (x = y)]$ la cual expresa llanamente que si una cosa es idéntica a otra, entonces son *necesariamente* la misma cosa.

La conclusión (4) parece demasiado estricta y ha suscitado, como el mismo Kripke lo señala, muchos argumentos en contra. No obstante, él sostiene que se da este tipo de identidad —i.e. una identidad necesaria— cuando las dos expresiones vinculadas en la relación de identidad son *designadores rígidos*—es decir, expresiones que designan una misma entidad en todos los mundos posibles en que ésta aparece— y no simples *descripciones definidas*. Así, por ejemplo, mientras el nombre propio Simón Bolívar es un designador rígido, pues designa o selecciona a la misma entidad en cualquier mundo posible donde aparece Simón Bolívar (dado que no podemos pensar en uno en el que Simón Bolívar no sea Simón Bolívar), “el libertador de Colombia” no lo es, puesto que siempre es posible imaginar un mundo donde Simón Bolívar no haya libertado a este país. La razón de ello es que “ser el liber-

tador de Colombia” no es una *propiedad esencial* de Simón Bolívar —aunque sí, una descripción definida de ese personaje en un determinado mundo posible—, pero “ser Simón Bolívar” sí lo es. Kripke sostenía, en consecuencia, que son necesarias las identidades que ponen en correspondencia designadores rígidos, i.e. expresiones que refieren a sus objetos en virtud de sus propiedades esenciales, de forma que ambos lados de la identidad habrían de seleccionar al mismo objeto en todos los mundos posibles. Esto elimina la ilusión de contingencia en muchas identidades y las convierte en verdades necesarias —si bien, puede que corroborables a posteriori—. De ahí que también se deba eliminar la ilusión de contingencia en la tesis de identidad de tipos, pues se sostiene que:

Si x es un dolor e y el estado cerebral correspondiente, entonces *ser un dolor* es una propiedad esencial de x , *y ser un estado cerebral* es una propiedad esencial de y . Si la relación de correspondencia es, de hecho, la identidad, entonces tiene que ser *necesario* de y que corresponda a un dolor y *necesario* de x , que corresponda a un estado cerebral; en realidad, a este estado particular y ” (11).

En este sentido, —afirma un poco más adelante— se dice que “mi dolor” y “que yo esté en estado cerebral x ” son designadores rígidos, pues “siempre que algo es tal y cual dolor, es esencialmente ese preciso objeto, a saber: tal y cual dolor; y siempre que algo es tal y cual estado cerebral, es esencialmente ese preciso objeto, a saber: tal y cual estado cerebral” (11). Esto significaría que, en caso de ser cierta la tesis de identidad, la relación que propondría entre el dolor y el estado cerebral tendría que ser necesaria y no contingente, como sus partidarios afirman, porque es esencial de la sensación de dolor el hecho de que sea sentida como dolor (“que una sensación sea *sentida* como dolor es que *sea* dolor”, (11), así como es esencial para el estado cerebral x el hecho de que sea tal estado cerebral. Pretender sostener una contingencia implicaría aceptar que el dolor que siento podría existir sin que yo experimentara dolor, lo cual es un absurdo. En consecuencia, Kripke termina concluyendo que cualquiera que defienda la teoría de identidad, no podría sostener, sino que debería negar, que podrían haber existido situaciones en las que uno hubiese tenido dolor, pero no el estado cerebral correspondiente (11). Pero esto va en contra de la teoría de identidad, por lo cual debe ser falsa².

² El argumento de Kripke es, por supuesto, mucho más complejo de lo que aquí se ha expuesto, y algunos autores consideran que es absolutamente concluyente. Otros, por su parte, sostienen que presenta algunas fallas, como a la hora de determinar cuáles son los mundos posibles o cuáles las propiedades esenciales de uno u otro designador rígido. Sin embargo, para efectos de este artículo, es suficiente con lo dicho hasta este punto.

Un segundo argumento que puede ser utilizado en contra de la teoría de identidad —y, en general, contra cualquier tipo de doctrina reduccionista— es esgrimido por Thomas Nagel. En éste se busca poner en evidencia que el reduccionismo no puede (ni podrá) dar cuenta de un rasgo esencialmente fenomenológico de la conciencia ni, en general, de la experiencia mental. “La razón —dice él— es que cada fenómeno subjetivo está esencialmente conectado con un solo punto de vista, y parece inevitable que una teoría objetiva, física, abandone ese punto de vista”. (12). Esa característica, absolutamente privada y dependiente del punto de vista del sujeto, se ha denominado el carácter *qualia* de los estados mentales. Para explicarla, Nagel propone la pregunta: “¿qué es ser como un murciélago?”, dado que, en general, tendemos a pensar que tener una experiencia consciente significa “básicamente, algo que es ser como ese organismo”. (12). En su ejemplo, nos dice que la ciencia ha demostrado que los murciélagos carecen de visión, pero que, poseen una suerte de sistema de sonar tan impresionantemente preciso que les permite percibir la distancia y configuración de los objetos de su entorno en virtud de la ecolocalización. La idea, en breve, es: aunque sepamos objetivamente cómo funciona esa ecolocalización, a tal punto que podemos imaginarnos cómo sería para nosotros percibir el mundo de

tal modo, no podríamos nunca saber exactamente cómo un murciélago lo percibe, pues para saberlo deberíamos *ser* murciélagos. Del mismo modo, no sabemos cómo puede ser la percepción de algo verde por otra persona, porque para hacerlo deberíamos *ser* esa persona, y no lo somos. Nuestros reportes respecto a cómo funcionan las percepciones sonoras de los murciélagos, e incluso las visuales de las demás personas, podrán ser todo lo objetivas que se quiera, pero su rasgo cualitativo —el carácter *qualia*— en cuanto vinculado íntimamente con el punto de vista del sujeto que percibe, es inexpresable por cualquier visión externa. De ahí que concluya que cualquier enfoque reduccionista, esto es, objetivo y postulado desde el punto de vista de la tercera persona, dejará siempre de lado el carácter subjetivo de la experiencia y, por lo tanto, será incompleto.

Smart habría vaticinado un contraargumento de este estilo, y propuso una posible respuesta para él a través de la inclusión de lo que denominó un *lenguaje cuasilógico o temáticamente neutral*. Su idea, palabras, más palabras menos, es que cuando una persona dice “tengo una postimagen naranja amarillosa” en realidad está diciendo algo como esto: “*Está pasando algo que es como lo que pasa cuando yo tengo mis ojos abiertos, estoy despierto y hay una naranja iluminada con*

buena luz frente a mí, esto es, cuando yo realmente veo una naranja” (6). De este modo, las palabras en *italica* deben interpretarse como una situación típica que es causalmente responsable del reporte de una experiencia de post-imágenes naranjas amarillosas. Así —ahora en términos de Place—, cuando un sujeto dice tener una experiencia mental x , lo que el psicólogo debe mostrar es “que el proceso cerebral que está causando que el sujeto describa su experiencia de tal modo, es el tipo de proceso que normalmente ocurre cuando...” (5), donde los puntos suspensivos hacen referencia a una situación típica que causa la experiencia mental x —en Lewis (10) se sostiene una idea equivalente—. Esto, por supuesto, no quiere decir que la experiencia mental no tenga propiedades; implica, en cambio, que “al hablar de ellas como siendo similares o disimilares una a otra no necesitamos conocer o mencionar esas propiedades” (6). Y tal cosa evita, por otra parte, que debamos pensar en algún tipo de lenguaje que exprese cualidades privadas de las experiencias.

No obstante, podría objetar un defensor de los *qualia*, uno podría imaginar a un neurofisiólogo que sepa todo acerca del cerebro, pero que, por alguna razón (v.gr., un defecto visual congénito), nunca haya tenido una “sensación-de-rojo”. En este caso, aunque dicho científico tenga un conocimiento objetivo de

todo lo que, ocurre en las neuronas cuando hay una sensación de rojo —y esto incluye desde las situaciones donde normalmente se produce, hasta aquellas cuando es extraño que lo haga—, habría algo que no sabría: “qué es tener una sensación-de-rojo” (8). La razón es sencilla y, de hecho, similar a la conclusión que dedujo Smart: no tenemos un lenguaje que nos permita hablar con objetividad de los *qualia*, pero esto no quiere decir que esas experiencias absolutamente subjetivas no sean *algo*. Que no podamos describirlas no quiere decir que no estén; simplemente, tiene que ver con que contamos con un sistema de representación lingüística que no cobija ciertos hechos subjetivos (12).

Finalmente, es importante referirse a una tercera crítica, que surgiría de una significativa formulación de Fodor, la cual ataca la estructura misma de la reducción interteórica. Este filósofo cree que el reduccionismo se apoya en “el supuesto de que [si] el tema de estudio de la psicología es parte del tema de estudio de la física [entonces ha de tomarse] como implicando que las teorías psicológicas deban ser reducidas a teorías físicas”, y esto último parece ser “lo que ocasiona el problema” (13). Ciertamente, como lo enunciaba más arriba, una de las ventajas que se desprendería del hecho de que fuera cierta la teoría de identidad es que siendo —como es— la física una ciencia de

categorías y términos precisos, una psicología expresada en su lenguaje nos proveería de categorías y términos igualmente precisos. Esta re-clasificación es, justamente, el objetivo final del reduccionismo positivista en psicología, pues alcanzarla significaría eliminar el ambiguo y casi inmanejable lenguaje intencional, a favor de uno que reflejase mejor la realidad, para permitirnos así predecir y explicar mejor los fenómenos que llamamos mentales. No obstante, para llegar de la ciencia por reducir a la ciencia reductora, debe pasarse por un proceso un poco más complicado de lo que a simple vista podría parecer. La reducción, a grandes rasgos, ocurre así (13): supóngase que la siguiente es una fórmula de la psicología:

$$(1) S_1x \rightarrow S_2y$$

donde S_1 es un predicado psicológico que se aplica a ciertos objetos, y x , y S_2 es otro predicado psicológico que se aplica a ciertos y . Para que esta ley pueda ser reducida a una ley física (3), cada uno de estos predicados psicológicos habrá de ser reducido, por su parte, a predicados físicos en una suerte de leyes llamadas *leyes puente* (2a) y (2b), así:

$$(2a) S_1x \leftrightarrow P_1x$$

$$(2b) S_2y \leftrightarrow P_2y$$

$$(3) P_1x \rightarrow P_2y$$

donde P_1 y P_2 son predicados de la física. El reduccionismo, desde la

perspectiva de la teoría de identidad de tipos, supone que las leyes puente deben interpretarse como identidades (contingentes) de tipos o clases de eventos, de forma que v.gr. (2a) se leería: todo tipo de eventos x que satisfaga S_1 es idéntico a algún tipo de eventos x que satisfaga P_1 , y viceversa. Esto implicaría, si es cierto, que toda clase de eventos psicológicos *es* una clase física (13). Sin embargo, no parece muy posible que la doble implicación de las leyes puente pueda mantenerse, pues es poco factible que cada clase psicológica corresponda a una clase física. Esto puede ser explicado del siguiente modo: supóngase que se quiere reducir una ley económica como “todo intercambio monetario justo implica una ganancia nula” a una ley física como (3). Según el modelo, esto sólo sería posible si pudiera establecerse una correlación uno-a-uno entre, primero, la clase de “los intercambios monetarios” y una clase física P_1 ; y, segundo, la clase de “ganancias nulas” y la clase física P_2 . Pero sería muy poco plausible que todos los intercambios monetarios, sin excepción, se correlacionaran con una u otra propiedad física: es bien sabido que pueden efectuarse con monedas, billetes, conchas y cualquier objeto al que se le estime algún valor. En este caso, la reducción no ayudaría en absoluto, pues lo que es común en los intercambios monetarios, es decir, la propiedad que los hace ser miembros de una misma clase, na-

da tiene que ver con los objetos físicos a través de los cuales se llevan a cabo. De forma análoga, “[s]i la psicología es reductible a la neurología, entonces para cada predicado de la clase psicológica hay un predicado de la clase neurológica, y la generalización que establezca esta coextensión es una ley” (13). Pero los datos con que cuenta la neurología para poder hacer sus correlaciones entre predicados psicológicos y cerebrales aún no lo permiten, y la falta de pruebas puede llevarnos a poner seriamente en duda la plausibilidad de una teoría reduccionista de tipos. ¿Cuál podría ser entonces la salida del reduccionismo, dadas esta y las anteriores objeciones a la teoría de identidad?

Materialismo eliminativo

Una manera de salir del atolladero donde se encuentra la teoría de identidad consistiría en una perspectiva que se inclinase a favor de un solo lado de la identidad; a saber: el de la neurofisiología. Esta visión fue originalmente sostenida, durante la década de los sesenta, por Richard Rorty y Paul Feyerabend, y actualmente la han retomado filósofos como Paul y Patricia Churchland, bajo el nombre de *materialismo eliminativo*. La idea central de esta postura es que la correlación uno-a-uno entre los estados mentales y los estados cerebrales, que resultaría de la reducción inter-teórica, es imposible, por

el simple hecho de que “el marco de nuestra psicología del sentido común es una concepción falsa y mal encaminada de las causas del comportamiento humano y de la naturaleza de la actividad cognitiva” (8). Sostienen, entonces, que el viejo marco conceptual deberá ser eliminado y reemplazado por uno más preciso: el de la neurociencia. Y dicha sustitución de lenguajes tendrá lugar cuando la neurociencia haya madurado lo suficiente para revelarnos la pobreza de nuestras concepciones y la superioridad de las suyas (8). De esta forma, podría verse al materialismo eliminativo como una teoría de la mente que *elimina* a la mente y propone en últimas que: *no existen los estados o procesos mentales, sólo podemos hablar de estados o procesos neurológicos (Tesis del materialismo eliminativo)*.

Los argumentos que estos teóricos esgrimen en pro de su doctrina pueden dividirse en dos clases: los “más” técnicos y los “menos” técnicos. Probablemente la argumentación “más técnica” del materialismo eliminativo proviene del mismo Feyerabend (14), quien se preocupa por demostrar que la teoría de identidad realmente no elimina al dualismo, sino que más bien lo presupone. La razón para ello —dice— es que una ley puente del estilo “ x es un proceso mental de la clase $A \leftrightarrow x$ es un proceso central [sic] de la clase a ”, parece implicar tanto que los eventos mentales tienen caracte-

rísticas físicas, como que los eventos físicos o cerebrales (que denominan centrales) tienen características mentales. Esto lo lleva a pensar, en consecuencia, que la formulación reduccionista de la teoría de identidad supone un dualismo de rasgos (o de propiedades), por lo cual resulta falsa la presunción de que ésta es una buena forma de refutar el dualismo. Lo máximo que logra hacer dicha doctrina —agrega— es redefinir los conceptos mentales, lo cual es una forma de perpetuarlos, y considera que lo mejor que podrían hacer sus partidarios sería desarrollar una teoría que no apelase a los recursos lingüísticos existentes. Cree, por lo tanto, que deberíamos abandonar el lenguaje mentalista y sustituirlo por el lenguaje de la neurociencia, del mismo modo en que se abandonó el lenguaje de las posesiones demoniacas una vez se desarrolló la teoría moderna de la epilepsia (14).

Los argumentos “menos” técnicos se basan, corrientemente, en este tipo de comparaciones históricas. Paul Churchland (8), por ejemplo, sostiene tres razones principales para preferir el enfoque del materialismo eliminativo. Primero, por las fallas explicativas, predictivas y de manipulación que, hasta ahora, ha demostrado tener la psicología popular. Segundo, porque en caso de ser cierto, se convertiría en una posición mucho más fuerte que las antes vistas. Y tercero, porque la histo-

ria de la ciencia nos ha demostrado que muchas teorías “problemáticas” del pasado han sido reemplazadas por otras más fuertes, con mayor poder explicativo y predictivo. La razón por la cual la psicología popular no ha sido eliminada, sugiere él, no es por el hecho de que sus representaciones sean correctas, sino más bien porque los fenómenos de los que se ocupa son tan “difíciles” que no se pueden eliminar muy fácilmente (8).

Estos argumentos, a todas luces, son muy poco contundentes, y considero que podrían objetarse fácilmente: en primer lugar, las fallas de la psicología popular podrán ser un motivo para dudar de ella, pero esto no significa que sean razones para confiar en el materialismo eliminativo. Por otra parte, apelar al hecho de que el materialismo eliminativo sería más fuerte que las demás doctrinas es una razón que aún está por verse, pues si bien en principio tal hipótesis parece verdadera, sólo en el momento en que contemos con las herramientas suficientes para verificarla podremos dar cuenta de ella. Además, es posible pensar que pese a que la neurociencia es una disciplina exacta y rigurosa, las explicaciones que pueda proveernos para reemplazar las psicológicas quizás no vayan a servirnos en la práctica. Y, finalmente, creo que las razones que aluden a la historia de la ciencia y de los “reemplazos teóricos” están franca-

mente mal pensadas. Dentro de las razones históricas, por ejemplo, es común que los materialistas eliminativos vaticinen un futuro para la psicología popular idéntico al que sufrió la teoría de la combustión del flogisto. Durante la Edad Media se pensaba que las cosas podían incendiarse en virtud de una cierta sustancia —el flogisto— que permitía la combustión; cuando la ciencia avanzó, el flogisto fue prontamente reemplazado por una explicación muy distinta, que involucraba sólo al conocido y ampliamente aceptado oxígeno. La ciencia, pues, se encargó de eliminar cualquier justificación para creer en la existencia del flogisto; similarmente —aseguran ellos—, la neurociencia se encargará de eliminar cualquier posible justificación para sostener las entidades que propone la psicología popular.

¿Cuáles fueron las razones que pudieron haber llevado a los científicos medievales a la postulación de tan extraña sustancia? Tengo la impresión de que a ellos les ocurrió algo semejante a lo acaecido más recientemente en la genética mendeliana. La historia, a grandes rasgos, es la siguiente: Mendel postuló, en 1865, la existencia de ciertos factores que consideraba responsables de la herencia de los rasgos entre padres e hijos. Sin embargo, él nunca pudo corroborar la existencia de tales fenómenos, pues fue sólo hasta que la ciencia descubrió los cromosomas cuando sus hipotéticos *genes* pudie-

ron ser verificados. Pues bien, creo que tanto Mendel como los científicos medievales manejaron un tipo de procedimiento (hipotético) científico similar: dedujeron que, dados ciertos fenómenos (combustión, en un caso, rasgos heredados, en el otro), debía de haber un cierto fenómeno x que pudiera ser su causa. Empero, la naturaleza de esta variable fue *sumamente bien delimitada*; es decir, la x realmente podía ser satisfecha por muy pocas cosas del mundo. Los medievales que postularon el flogisto se equivocaron; Mendel postuló los genes y, *voilà*, acertó. Pero, realmente, era casi más difícil equivocarse que atinar: las características que debía tener el flogisto para comportarse como la causa de la combustión estaban bastante bien delimitadas; el problema es que podían ser satisfechas también por el oxígeno, y este segundo elemento resultó ser, efectivamente, el responsable de las igniciones. El asunto fue, podría decirse, de mala suerte. ¿Qué sucede, en consecuencia, con el materialismo eliminativo? El problema está en que para ser cierto, aquello que llamamos mental y que postulamos como responsable de ciertos eventos comportamentales debería, al igual que las x de la conjeturas medievales y mendelianas, estar bastante bien delimitado. Pero esto no es el caso. No existe una definición unánime, por ejemplo, entre lo que es una actividad cognitiva y lo que no lo es. El concepto de cognición parece abarcar desde las operaciones requeridas

para oprimir un botón del control remoto, hasta las que se necesitan para conducir en el espacio un transbordador espacial. Y ni qué decir de las sensaciones, los recuerdos, los deseos e, incluso, aquellas cosas que denominamos con términos más ordinarios, como estados de ánimo, neuras, manías, etc. Simplemente, no existe un límite claro que defina hasta dónde van las actividades mentales. Y sin ese límite no hay siquiera una idea remota sobre cuál es el “flogisto mental” que debe reemplazarse por el “oxígeno neurológico”.

El materialismo eliminativo, en todo caso, ha recibido muchas críticas durante los últimos años. Algunos sostienen que los objetivos del materialismo eliminativo difieren de los de la psicología popular por el hecho de que ésta no es un proyecto de investigación. Otros, por su parte, consideran que en realidad la teoría eliminativista no es una demostración que busque dejar de lado al mentalismo, pues parece que “desde el principio, niega la existencia de algo a lo que hay que dejar de lado” (15). Pero quizás una de las más interesantes objeciones —que llega a una conclusión similar a la que antes habría expuesto Fodor en contra del reduccionismo—, es planteada por Searle (15) a través de un ejemplo (el cual he abreviado aquí ligeramente) que busca reflejar el absurdo al que se llega con el tipo de argumentación que sostiene el

materialismo eliminativo. Considérese, pues, la estructura del siguiente argumento a favor del materialismo eliminativo:

Argumento 1:

1. Imaginemos que tuviéramos una ciencia neurobiológica que explicara perfectamente cómo funciona nuestro cerebro.
2. Tal ciencia cubriría el mismo campo que la psicología popular, pero sería más poderosa.
3. Además, parece altamente improbable que los conceptos ordinarios de la psicología popular (v.gr. “creencia” o “deseo”), casaran con los de la taxonomía de esta neurociencia.
4. Esta neurociencia, entonces, no deja lugar para conceptos como creencia o deseo, y no sería posible reducirlos adecuadamente.
5. *Por lo tanto*, las entidades nombradas por la psicología popular no existen.

Ahora compárese con este otro:

Argumento 2:

1. Consideremos la ciencia física.
2. Esta teoría explica perfectamente cómo funciona la realidad física y resulta ser enormemente superior a nuestras teorías del sentido común.
3. Además, nuestros conceptos ordinarios de la física del sentido común (v.gr. club de golf o casa de campo) no casan, ni exacta ni remotamente, con la taxonomía de la física.

4. No hay, entonces, un uso en la física teórica para estas expresiones, pues la manera en que ésta taxonomiza la realidad es distinta del modo en que lo hace la física popular, y no habría lugar para un reducción adecuada.
5. *Por lo tanto*, ni los clubes de golf ni las casas de campo existen realmente.

El problema del materialismo eliminativo, concluye Searle, es que:

...descansa en la premisa, obviamente falsa, de que, para cada teoría empírica y su correspondiente taxonomía, a no ser que haya una reducción de tipo a tipo de las entidades taxonomizadas a las entidades de las mejores teorías de la ciencia básica, las entidades no existen (15).

Aunque el argumento tal vez parezca un poco trillado, tengo la impresión de que apunta a la verdad de las cosas. En todo caso, es suficiente dejar el asunto así por lo pronto, pues diré algo más al respecto en la última sección.

Bibliografía

1. Putnam H. The nature of mental states. In: Block N. Readings in philosophy of psychology. Cambridge: Harvard University Press; 1980. p. 223-31.
2. Ryle G. The concept of mind. London: Hutchinson; 1949.
3. Putnam H. Brains and behavior. In: Butler RJ, editor. Analytical philosophy. Oxford: Blackwell; 1965.
4. Chomsky N. A review of B. F. Skinner's verbal behavior. *Language*;35(1):26-58.
5. Place UT. Is consciousness a brain process? In: Lycan WG, editor. Mind and cognition. Cambridge: Blackwell; 1990.
6. Smart JJC. Sensations and brain processes. In: Chappell VC, editor. The philosophy of mind. Englewood: Prentice Hall; 1962.
7. Wittgenstein L. Investigaciones filosóficas. Barcelona: UNAM-Crítica; 1988.
8. Churchland PM. Matter and consciousness. Cambridge: MIT Press; 1988.
9. Frege G. Sobre sentido y referencia. In: Valdés JM, editor. La búsqueda del significado. Madrid: Tecnos; 2000.
10. Lewis DK. Un argumento en favor de la teoría de identidad. *Cuadernos de Crítica* 1984;30.
11. Kripke S. Identidad y necesidad. In: Valdés JM, editor. La búsqueda del significado. Madrid: Tecnos; 2000.
12. Nagel T. What is like to be a bat? In: Block N. Readings in philosophy of psychology. Cambridge: Harvard University Press; 1980. p. 159-68.
13. Fodor JA. Special sciences. In: Fodor JA. Representations. Cambridge: MIT Press; 1981.
14. Feyerabend PK. Mental events and the brain. In: Lycan WG, editor. Mind and cognition. Cambridge: Blackwell; 1990.
15. Searle JR. El redescubrimiento de la mente. Barcelona: Crítica; 1996.