

UN ANALISIS LOGICO DE LAS TEORIAS DE LA IDENTIDAD PSICONEURAL† (*A logical analysis of the psychoneural identity theories*)

María José FRAPOLLI*

Manuscrito recibido: 1999.5.6.

Versión final: 2000.3.2.

* Departamento de Filosofía, Facultad de Psicología, Universidad de Granada, 18071 Granada. E-mail: frapolli@platon.ugr.es

BIBLID [0495-4548 (2000) 15: 38; p. 319-348]

RESUMEN: El propósito de este artículo es mostrar que las llamadas "teorías de la identidad psiconeural" no tienen la estructura lógica de una teoría genuina de la identidad. Un operador de identidad genuino es un operador de segundo orden, o de órdenes superiores, que convierte predicados n -ádicos en predicados $n-1$ -ádicos. Si las teorías de la identidad psiconeural no son, en realidad, teorías de la *identidad*, entonces la crítica habitual que usan la Ley de Leibniz y otros rasgos que se relacionan con la identidad no tienen ningún efecto. Así, dos serán las conclusiones: (i) que las llamadas "teorías de la identidad psiconeural" deben de reformularse, de manera que expresen claramente su contenido, y (ii) que las críticas utilizadas hasta ahora contra ellas no consiguen su objetivo.

Descriptores: identidad tipo-tipo, Ley de Leibniz, operador de identidad, eventos, descripciones de eventos.

ABSTRACT: *The aim of this paper is to show that the so-called "theories of psychoneural identity" do not possess the logical structure of a genuine theory of identity. The identity operator is a second- or higher-order operator that converts n -adic predicates into $n-1$ -adic ones. If the so-called "theories of psychoneural identity" are not theories of identity at all, then two conclusions follow: (i) these theories of the mind-body relation should state their contents in a more appropriate way, and (ii) the usual criticisms against them, which bear on some features that all identity theories must hold, as for instance Leibniz Law, leave them untouched.*

Keywords: *Type-type identity, Leibniz Law, identity operator, events, descriptions of events.*

SUMARIO

1. Introducción: Monismos y Dualismos

1.1. La Ley de Leibniz

2. Teorías de la Identidad Psiconeural

3. Críticas al núcleo central de la TIP

3.1. Los argumentos de la diversidad de propiedades

3.1.1. Los argumentos epistemológicos

3.1.2. El argumento de la localización espacial

3.1.3. El argumento de las propiedades fenoménicas

3.2. El argumento de la realizabilidad variable

- 3.3. El argumento lógico-semántico de Kripke
4. El operador de identidad
5. La identidad psiconeural y el relator de identidad
6. Conclusión

1. Introducción: Monismos y Dualismos

¿Qué es el hombre? Pocas cuestiones filosóficas son más antiguas y más controvertidas que ésta que pregunta por la naturaleza humana. Pocas también habrán merecido una mayor cantidad de repuestas y un repertorio de respuestas que abarquen un espectro más amplio. A pesar de toda su dificultad, sin embargo, las teorías acerca de lo que es el hombre pueden clasificarse en dos grandes grupos: las que afirman que el hombre es un ser compuesto o híbrido, esencialmente diferente del resto del mundo natural, y las que afirman que el ser humano forma parte de la naturaleza como cualquier otro animal, sin que haya nada en él que lo haga esencialmente distinto del resto de las criaturas ni merecedor de una categoría aparte. La diferencia radical que el primer grupo de teorías encuentra en el ser humano suele tener que ver con la postulación de una mente, de un alma, o de un yo, en cualquier caso una entidad no física, que permitiría explicar la complejidad de la actividad humana, que es a veces una actividad libre, o que está motivada por convicciones y principios, o que va dirigida a fines, o que es objeto de evaluación moral, etc. El segundo grupo de teorías no propone una imagen del ser humano muy diferente en cuanto a su actividad, pero las teorías reunidas en él parten del supuesto de que esa enorme complejidad puede explicarse en principio sin postular entidades supranaturales. O, en versiones más débiles, que si no puede explicarse sin apelar a sustancias especiales, entonces no puede explicarse en absoluto. Si bien es cierto que la situación actual de la ciencia no puede dar razón de por qué cuando me doy cuenta de que llego tarde, decido tomar un taxi en vez de proseguir con mi agradable paseo por la ciudad, las teorías del segundo grupo suponen que la explicación de esta conexión o está en la ciencia (presente o futura) o no está en ningún lado.

A las teorías del primer grupo se las conoce como Teorías Dualistas, teorías que afirman que el hombre es un compuesto de dos sustancias, y a las teorías del segundo grupo se las conoce como Teorías Monistas¹, Materialistas o Fisicalistas.

Las Teorías Dualistas analizan al ser humano como si fuera un compuesto de dos tipos de sustancias², una sustancia material y una sustancia no-

material (mental o espiritual), con comportamientos y propiedades distintas. Los dos antecedentes históricos más famosos son Platón y Descartes, aunque es éste último el reponsable de la forma particular que en la actualidad toman las teorías dualistas y de los argumentos a favor y en contra de las mismas. Descartes tiene el mérito añadido de ser el primer filósofo que se planteó con claridad la mayor dificultad filosófica con la que se enfrenta todo dualismo de sustancias, a saber, el problema de explicar la interacción causal entre la materia y el espíritu. Parece claro que esa interacción ha de producirse de algún modo³, porque solemos explicar nuestra conducta conectando causalmente deseos, intereses, objetivos, etc. con nuestra particular forma de actuar. Acostumbramos a decir que fue su deseo de quedar bien con sus invitados lo que le llevó a encargar la cena a la mejor empresa de *catering*, o que porque tenía interés en conseguir la beca, renunció a sus vacaciones los últimos años y se dedicó a trabajar.

El dualismo de sustancias no sólo tiene un interés histórico y en la actualidad ha sido reelaborado y defendido por Popper y Eccles, por Swinburne y por Foster, entre otros⁴.

El Monismo Materialista o Fisicalista defiende la tesis de que nuestra vida mental es una expresión de nuestra actividad cerebral, que los seres humanos formamos parte del mundo natural y que no hay más vía para entender lo que somos que las que nos proporciona la ciencia presente o futura. La versión del Monismo Materialista que se analizará en lo que sigue es la denominada "Teoría de la Identidad Psiconeural" y su tesis definitoria se resume en la afirmación de que los estados mentales no son más que estados cerebrales o neuronales.

En la filosofía de la mente contemporánea se ha abandonado casi por completo el discurso cartesiano que hablaba de sustancias, en favor de un discurso que habla de estados mentales y cerebrales o, alternativamente, de eventos o procesos mentales y cerebrales. Aquí hablaremos indistintamente de estados o de eventos, puesto que para la tesis que se defenderá en lo que sigue es irrelevante una opción u otra.

Bajo el rótulo general de "Teorías de la Identidad Psiconeural" se han refugiado teorías de categorías muy diversas. En líneas generales y sin ninguna pretensión de exhaustividad, tres suelen ser las categorías en las que las teorías de la identidad se clasifican: Teorías de la Identidad del Tipo, Teorías de la Identidad de las Instancias y Teorías del Doble Aspecto.

Las Teorías de la Identidad del Tipo afirman que los estados mentales son tipos de estados neuronales. Al hacer la identificación entre tipos de

estados mentales y neuronales, estas teorías de la identidad permiten el establecimiento de leyes generales. Las Teorías de la Identidad de las Instancias afirma que todo estado mental es idéntico a alguna configuración cerebral, no necesariamente siempre la misma, esto es, que los estados mentales poseen siempre algún soporte cerebral, con lo que la posibilidad de establecer regularidades nomológicas se desvanece. Esta versión es mucho más débil que la anterior y es compatible con algunos tipos de dualismo. Por último, las Teorías de la Identidad del Doble Aspecto postulan un sustrato neutral, ni físico ni mental, del cual las propiedades mentales y físicas serían manifestaciones. Esta versión, históricamente relacionada con Spinoza, es compatible (y a veces se identifica) con el llamado "Dualismo de Propiedades"⁵.

En el presente trabajo nos centraremos en las teorías de la identidad del tipo, la versión más fuerte. Mostraremos cómo estas teorías se proponen a sí mismas como teorías de la identidad, y cómo la mayoría de sus detractores las critican precisamente por no cumplir con algunos de los principios que definen esta noción. No obstante, mostraremos que el operador de identidad necesario para exponer enunciados de identidad genuinos no puede ser el habitual operador diádico de primer orden, representado por el signo "=", sino un operador de un orden superior que funciona como una especie de cuantificador. Este nuevo operador de identidad plantea un problema a las llamadas "teorías de la identidad", puesto que sus tesis principales no pueden reformularse como enunciados de identidad genuinos en los que aparezca el operador nuevo. Y esto significa que las teorías que explican la relación mente-cerebro como una relación de identidad están equivocadas en cuanto a la forma lógica de su propia teoría. No pretendemos afirmar que lo que las teorías de la identidad mente-cerebro dicen sea falso, pero sí que su contenido debería de expresarse de alguna otra forma más adecuada. Esta es la conclusión negativa del artículo. La conclusión positiva sería que, si la interpretación que se defenderá en lo que sigue del operador de identidad es correcta, las teorías de la identidad psiconeural se vuelven inmunes a las críticas habituales que utilizan la Ley de Leibniz y otras características de la noción de identidad.

1.1. *La Ley de Leibniz*

La teoría de la identidad psiconeural (TIP, en adelante), al proponerse como una teoría de la *identidad*, se rige por los estándares lógicos que habitualmente se relacionan con esta noción: en primer lugar, lo que habi-

tualmente se conoce como "la Ley de Leibniz" (LL) y, en segundo lugar, los axiomas que expresan las propiedades que hacen de la identidad una relación de equivalencia. Mediante la expresión "Ley de Leibniz", los filósofos se refieren a principios muy diversos y no está claro cuál es la posición exacta del Leibniz histórico. Es útil en este contexto distinguir entre

- (LL1) la indiscernibilidad de los idénticos, que expresa una condición necesaria de la identidad,

$$(x) (y) (x=y \rightarrow (Px \leftrightarrow Py)),$$

- (LL2) la identidad de los indiscernibles, que expresa una condición suficiente de la identidad,

$$(x) (y) ((Px \leftrightarrow Py) \rightarrow x=y),$$

- (LL3) el principio de sustitutividad: dos expresiones co-referenciales pueden sustituirse en todos los contextos *salva veritate*. (Este es un principio metalingüístico que puede considerarse como una versión en el modo formal de la conjunción de (LL1) y (LL2)),

y por último

- (LL4) un principio de segundo orden que afirmarí­a identidad entre conceptos⁶.

La teoría de la identidad parte de (LL1), (LL2) y (LL3) y añade además los siguientes tres axiomas:

(Reflexividad) $(x) (x=x)$

(Simetría) $(x) (y) (x=y \rightarrow y=x)$

(Transitividad) $(x) (y) (z) ((x=y \wedge y=z) \rightarrow x=z)$

La mayor parte de las críticas contra la teoría de la identidad psiconeural la acusan de violar uno o varios de los principios presentados en esta sección.

2. Teorías de la Identidad Psiconeural

La TIP, tal como se conoce en la actualidad, se localiza históricamente en la segunda mitad de los años cincuenta y en la década de los sesenta. Sus defensores más representativos fueron U.T. Place, J.J. Smart y H. Feigl⁷.

La teoría deriva (y es una mejora) de la posición conductista en psicología, iniciada por J.B. Watson y B.F. Skinner⁸, y utiliza así mismo aspectos del conductismo lógico defendido por G. Ryle⁹, entre otros. El avance de la TIP sobre el conductismo consiste en el reconocimiento de procesos y estados internos como causantes de la conducta. El conductismo explicaba los estados epistémicos ("creo que Pierre Menard es el autor del Quijote", "sé que el bardo Ossian es James Mc.Pherson") y volitivos ("me gustaría convertirte en un escarabajo de la patata", "deseo reinventar la humanidad") como si fueran propiedades disposicionales. Esta explicación fue aceptada por los representantes de la teoría de la identidad quienes, sin embargo, la consideraron inadecuada para dar cuenta tanto de las sensaciones (esto es, de las peculiares propiedades fenoménicas de la visión del rojo o de las especiales características que acompañan al estado de estar oliendo el perfume *Trésor*) como de la conciencia (esto es, de los estados en los que uno se percibe de su situación en el mundo y de sus relaciones con él). Todos ellos suponen que hay en el percibirse del estado actual de cada cual, en el darse cuenta del presente, y en el aspecto cualitativo que acompaña a las sensaciones, algo esencialmente interno y privado, que no puede reducirse a respuestas conductuales, ni siquiera potenciales. La inadecuada explicación de las sensaciones y la conciencia ha sido la razón que ha provocado el paso desde el conductismo a las primeras versiones de la TIP.

Armstrong¹⁰, otro de los proponentes clásicos de la teoría, presenta una versión de la misma diferente de la ofrecida por Place y Smart que posteriormente ha tenido mucha repercusión. Armstrong argumenta en favor de su posición que la ciencia física es la mejor explicación disponible de la realidad, de toda la realidad, y considera al hombre únicamente como un compuesto físico-químico y, por tanto, parte del mundo natural. Esta versión de la teoría opta, como las anteriores, por una concepción materialista o fisicalista del ser humano, aunque su materialismo no se presenta como una consecuencia lógica de la misma. A la versión de Armstrong de la teoría de la identidad se la denomina, a veces, la "Teoría de la Identidad del Rol Causal" porque hace un análisis de los procesos conscientes atendiendo al tipo de función que realizan y al tipo de relaciones causales en las que participan. A veces también se la conoce como "Monismo Neutral", porque propone una teoría de la identidad del doble aspecto.

Armstrong divide la explicación materialista de los procesos mentales en dos pasos; el primero consistente en la reformulación de los términos mentales y neuronales en una terminología neutral (ni mental ni física) y el

segundo en la afirmación contingente de la identidad de los procesos mentales con los procesos cerebrales.

La TIP, como se ha visto, ha sido defendida en múltiples versiones, pero su núcleo mínimo tiene como ingredientes fundamentales las dos tesis siguientes:

- (2.1) los procesos (eventos, estados) mentales no son, en último extremo, más que procesos (eventos, estados) cerebrales, y
- (2.2) la afirmación que se recoge en (2.1) es una hipótesis científica.

La tesis (2.1) equivale a un principio de parsimonia ontológica: si hacemos un recuento de los procesos (estados, eventos) básicos del universo no encontramos procesos cerebrales junto con procesos mentales de una clase distinta. Una cosa es cómo sentimos ciertos procesos o cómo los describimos y otra es lo que realmente hay. Suponer, a partir de las distintas formas de acceso a la realidad, tipos diferentes de existencia, cada uno correspondiente a una forma peculiar de conocer, es lo que Place denomina "la falacia fenomenológica", que él formula de la siguiente manera:

El error lógico, al que me refiriré como la falacia fenomenológica, es el error de suponer que cuando el sujeto describe su experiencia, cuando describe cómo le parecen las cosas, cómo suenan, saben, huelen o las siente, está describiendo las propiedades literales de objetos o eventos en una clase peculiar de cine interno, o pantalla de televisión, al que habitualmente se hace referencia en la literatura psicológica moderna como el campo fenoménico. (Place 1956, p. 49)

Hay, así, en la teoría una diferenciación entre los mecanismos epistemológicos que nos permiten conocer la realidad, por un lado, y la realidad misma, por otro.

Los retos más serios a la TIP, las cuestiones que más se han resitado a ser tratadas por la teoría, son precisamente aquellas en las que la diferenciación entre epistemología y metafísica es más difícil de mantener, a saber, en el problema de las propiedades fenoménicas, los *qualia*, y en el problema de la conciencia.

La tesis (2.2), a saber, que la identificación de los procesos mentales con procesos neuronales es una hipótesis científica, equivale a decir que las proposiciones en las que se expresa la teoría son contingentes y que el acceso que tenemos a ellas es *a posteriori*. Esta tesis se opone a la afirmación de que la TIP sea una teoría lógico-semántica¹¹, esto es, sus proponentes no afirman que las oraciones del tipo

- (2.2.1) el dolor es la estimulación de las fibras C

sean analíticas; que los dos términos que aparecen en la aparente oración de identidad, "el dolor" y "la estimulación de las fibras C", sean sinónimos; ni siquiera que sean reductibles uno al otro. Así, la TIP no supone que el lenguaje en el que expresamos nuestras experiencias perceptivas o conscientes sea, ni en la actualidad ni con el tiempo, sustituible por el lenguaje de la ciencia física. Esta última posición se ha mantenido de hecho, pero es diferente y más fuerte que la TIP. A la posición que sostiene que el lenguaje que usamos para hablar de los procesos mentales debería ser sustituido por un lenguaje en el que sólo aparecieran expresiones de las ciencias físicas se la conoce como "Materialismo Eliminatorio" y ha sido defendido, entre otros, por Feyerabend, los Churchland, Stich y Rorty.

La TIP supone que las oraciones como (2.2.1) son hipótesis de la ciencia, que pueden ser verdaderas o falsas, y que su contenido es empírico y no lógico. No son afirmaciones metateóricas. Place (1956, pp. 43-5) explica la confusión entre afirmaciones empíricas y lógicas en el caso de las identificaciones teóricas apelando a tres usos distintos del verbo "ser". Podemos usar "es" en el contexto de una definición:

(2.2.2) un oftalmólogo es un médico de ojos,

o bien para expresar "composición", como él lo denomina,

(2.2.3) una nube es una masa de pequeñas partículas en suspensión,

o bien para expresar predicación,

(2.2.4) un automóvil es un medio de locomoción.

En los dos primeros casos, la oración puede reformularse usando la expresión "no es más que". Así, podemos decir que un oftalmólogo no es más que un médico de ojos, o que una nube no es más que una masa de partículas en suspensión. Sin embargo, en el primer caso estamos ante una oración analítica de contenido lingüístico, mientras que en el segundo caso estamos ante oraciones empíricas cuya verdad o falsedad es sólo descubrible tras la investigación. La tesis (2.2) de la TIP viene a decir que las afirmaciones de la teoría son del tipo de (2.2.3).

La TIP ha sido objeto de numerosas críticas, aunque su inequívoco posicionamiento en el ámbito de la ciencia natural sigue ejerciendo un poderoso atractivo para muchos filósofos. A continuación veremos algunas de las críticas más frecuentes, aquellas que atacan a la teoría por ser, precisamen-

te, una teoría de la *identidad* y que han convencido a muchos filósofos de corte materilista de la conveniencia de buscar opciones naturalistas mejor protegidas de los ataques lógicos.

3. *Críticas al núcleo central de la TIP*

Las críticas a la teoría han sido muy diversas pero la gran mayoría se centran en los dos puntos más débiles de la misma: la explicación de los *qualia* y la explicación de la conciencia.

Para ilustrar el problema que los *qualia*, el aspecto fenoménico de la percepción, plantea a la TIP recurriremos a un famoso experimento mental, aunque muy posterior a las primeras versiones de la teoría que se han expuesto aquí. El experimento se debe a Frank Jackson¹² y consiste en lo siguiente: Mary, una brillante especialista en neurofisiología de la visión, ha vivido toda su vida en una habitación en blanco y negro y todo lo que conoce del mundo exterior lo conoce a través del monitor en blanco y negro de una pantalla de televisión. Por otro lado, tiene una excelente preparación en su campo y sabe todo lo que se puede saber acerca de la visión del color. En un momento determinado, Mary sale de su encierro y ve por primera vez el auténtico color de las cosas, el azul del cielo en verano, el rojo de las fresas, el verde de la hierba, etc. La cuestión que aquí se plantea es si Mary aprende algo nuevo del color, o de nuestra visión de color, tras su encuentro con el mundo real.

Jackson utiliza esta historia como contraejemplo a la identificación de las sensaciones con los procesos físicos mediante los que las explicamos en nuestras teorías científicas. Pero nuestro interés en esta historia es más modesto. Se trata en este caso de considerar dos vías diferentes de acercamiento al fenómeno de la visión de los colores: el conocimiento de los procesos mentales mediante los que captamos las longitudes de onda que nos envía la luz y nuestras sensaciones cuando de hecho los percibimos. Dos posiciones son en este caso posibles¹³. La primera, identificar los estados perceptuales internos con sus procesos neuronales y responder a la dificultad planteada por Jackson diciendo que Mary no aprende nada que no supiera ya y, la segunda, aceptar la parte perceptiva de la experiencia de Mary como signo de la imposibilidad de la identificación. En la primera posición, Mary reconoce por una vía distinta lo que ya sabía (que la hierba es verde, los tomates rojos, etc.). En la segunda posición, Mary apprehendería una nueva realidad (cómo se percibe de hecho el rojo de las fresas, el verde de la hierba, etc.). En este último caso tenemos ya un contraejemplo

a la TIP. En ambas opciones caben matizaciones. En la primera, puede pensarse que Mary tiene acceso a una única realidad de dos maneras diferentes. Esta sería la versión que se desprende del texto de Place citado anteriormente. En la segunda, se puede pensar que la parte perceptiva de la experiencia, lo que los filósofos en la actualidad llaman "*qualia*", es irreductible a los procesos neurofisiológicos pero que, sin embargo, no juega ningún papel causal en las relaciones de los sujetos con su medio. Esta posición se conoce como "Epifenomenalismo". O, finalmente, puede sostenerse así mismo que los *qualia* juegan algún papel funcional o causal aunque de naturaleza desconocida. Estas dos últimas interpretaciones supondrían una refutación de la teoría de la identidad, porque asumen que hay algo más en nuestros estados y procesos mentales que lo que puede expresarse atendiendo a los procesos del cerebro, y en este último caso estaríamos ante una posición dualista.

El problema de la conciencia es similar al problema de la percepción de lo *qualia*. De hecho, lo que aquí hemos llamado el problema de la percepción de los "*qualia*" se conoce a veces como el problema de la "conciencia fenoménica". Otras nociones de conciencia que aparecen en las discusiones sobre esta cuestión son la auto-conciencia, la conciencia como un pensamiento de orden superior acerca de nuestros estados mentales y la conciencia-acceso. A nuestro modo de ver, son las dos primeras nociones las que plantean el reto a la TIP. La autoconciencia es la capacidad que poseen algunos seres de atribuir a un yo, de atribuirse, estados y contenidos mentales. Me doy cuenta de que soy yo la que estoy viendo ahora la pantalla del ordenador, y sé que mi dolor de muelas es *mi* dolor de muelas. Sé que soy yo la que añora las últimas vacaciones y yo la que cree que eso que tengo ante mí es agua. También se entiende la conciencia como un estado cuyos contenidos son otros estados mentales, un estado capaz de examinar estados mentales de orden inferior. Soy consciente de que creo que hoy está siendo un día muy frío, me doy cuenta de que me preocupa cómo resultará este artículo, y estoy segura de que nadie sabe mejor que yo que estoy pensando en este momento que la cena está prácticamente lista. La dificultad que la conciencia (fenoménica o no) supone para la TIP consiste en la intuición de que una cosa son mis estados de creencia, deseo o dudas y otra distinta mi percibirme de todo ello. Y sin embargo no parece que el percibirse requiera un estado *neuronal* distinto de los estado neuronales que son su objeto, aunque sí parece un estado *mental* distinto. Para ilustrar esta duplicidad entre estados mentales distintos sobre un mismo estado físico se suele recurrir en estos casos a la posibilidad de la existencia de

zombies, seres sin autoconciencia pero con nuestras mismas respuestas y procesos, o a la posibilidad de construir artefactos capaces de poseer estados mentales pero sin reflexividad. La ciencia-ficción ha utilizado con profusión todas estas posibilidades. El análisis correcto tanto de los *qualia* como de la conciencia plantea dificultades para cualquier fisicalismo, también para las posiciones funcionalistas. Estrictamente hablando, la TIP exigiría que el sustrato físico que se identifica con las propiedades fenoménicas o con la conciencia fuera exactamente igual al cerebro humano. Si la identificación se produce con algún sustrato material no idéntico al cerebro estamos ante una afirmación de corte materialista o naturalista, pero no ante una versión de la TIP.

A pesar de que no puede hablarse de una única teoría de la identidad mente-cuerpo, sino de múltiples versiones, a veces muy diferentes entre sí en los detalles, todas las teorías de la identidad comparten un núcleo central, que hemos resumido anteriormente en las tesis (2.1) y (2.2), por lo que es posible exponer y evaluar las críticas a la TIP (o al menos algunas de ellas) sin necesidad de entrar en los pormenores que distinguen unas formulaciones de otras. Recordemos, no obstante, que estamos tratando aquí con teorías de la identidad del tipo, ya que las teorías de la identidad de las instancias son tan débiles que son incluso compatibles con posiciones dualistas.

Podríamos distinguir, de manera preliminar, dos grupos de constelaciones de críticas o, mejor aún, dos perspectivas, motivaciones o concepciones básicas desde las que se dirigen las críticas a las teorías de la identidad. Puede hablarse así de críticas desde el dualismo y de críticas desde lo que podríamos llamar, con un término amplio, el naturalismo. Las críticas desde el dualismo van encaminadas a mostrar que, apelando exclusivamente a los procesos neurofisiológicos, no puede explicarse la naturaleza humana y, en particular, el problema de la conciencia (en esta posición se encuentran Descartes, Popper y Eccles, y Kripke). Los críticos desde el naturalismo están interesados en mostrar que la solución fisicalista de las teorías de la identidad no es completamente adecuada, aunque sin rechazar el paradigma cientista (en este grupo se clasificarían concepciones como el funcionalismo, el emergentismo o las tesis de la superveniencia.)

Todas las críticas, no obstante, recurren en último extremo a los rasgos de la TIP que la definen como una *teoría de la identidad*, y en prácticamente todos los casos una de las armas utilizadas es la Ley de Leibniz, de la que ya hemos tratado. Dividiremos las críticas en tres grandes bloques:

- 3.1. Los argumentos de la diversidad de propiedades
- 3.2. El argumento de la realizabilidad variable
- 3.3. El argumento lógico-semántico de Kripke

Todas ellas pueden usarse en apoyo del dualismo, y las dos primeras se usan también a veces en favor de otras posiciones naturalistas. No entraremos aquí en la cuestión de si los argumentos esgrimidos contra la TIP son o no concluyentes, lo que nos interesa más bien es resaltar que, de hecho, se utilizan contra las teorías de la identidad y que todos ellos se centran en la (supuesta) estructura lógica de estas teorías. Veremos estos argumentos brevemente.

3.1. Los argumentos de la diversidad de propiedades

La estructura básica de todos los argumentos que sacan partido de la diferencia entre el tipo de propiedades que poseen los procesos mentales y los procesos físicos (los argumentos de la diversidad de propiedades) queda claramente representada en las disquisiciones cartesianas.

Yo puedo imaginarme a mí misma sin multitud de propiedades que poseo pero no puedo representarme sin mis pensamientos o sin mi autoconciencia. Puedo dudar de que esté ahora mismo ante el ordenador, escribiendo estas líneas. Todo lo que me rodea, mis seres queridos, mi casa, esta habitación, mi propio cuerpo y las sensaciones relacionadas con él pudieran ser el producto de un gran engaño. No puedo dudar, sin embargo, de que dudo, de que me pregunto si lo que veo, huelo o percibo por cualquiera de los sentidos externos es o no real, aunque pueda dudar de lo que veo, huelo o percibo. De aquí deduzco que soy algo que piensa, aunque no necesariamente yo soy sólo algo que piensa¹⁴. Si puedo dudar de mi cuerpo y puedo concebir, clara y distintamente, mi propia existencia sin él, pero no puedo dudar de ciertos actos de mi conciencia, entonces mi conciencia, mi yo, mi mente, mi alma o como queramos decirlo, no pueden identificarse con mi cuerpo, ni con mi cerebro ni con mi glándula pineal ni con ninguna otra parte de mi soporte físico y de este modo argumenta Descartes¹⁵ que mi mente es distinta de mi cuerpo porque puedo concebir clara y distintamente a la una sin el otro. Mi mente y mi cuerpo tienen propiedades básicas distintas: los objetos físicos, mi cuerpo entre ellos, tienen extensión mientras que la mente es inextensa. De ahí se sigue, en el argumento cartesiano, que cuerpo y mente, al poseer propiedades distintas, tienen que ser sustancias distintas. La solución cartesiana es, así, un dualis-

mo de sustancias. Con unos u otros matices, todos los argumentos de esta sección 3.1 tienen la misma estructura¹⁶. En las tres subsecciones que siguen se verán ejemplos de argumentos del tipo general de la diversidad de propiedades.

3.1.1. *Los argumentos epistemológicos*

Supongamos que la química contemporánea sea correcta y que el agua sea de hecho H₂O. Yo sé que el agua es H₂O pero mi hija de seis años no lo sabe, ella sabe que el agua quita la sed, pero no sabe que el H₂O quita la sed, luego el agua y el H₂O tienen propiedades distintas que tienen que ver con las diferentes actitudes epistémicas que los sujetos tienen hacia ellas. Lo mismo ocurre con mi dolor de cabeza en este momento y la excitación de ciertas fibras de mi cerebro. Es seguro que la humanidad ha sufrido dolor desde siempre y los seres humanos han aprendido muchas cosas acerca del dolor a lo largo de los siglos. Pero sólo desde hace unos pocos años se sabe algo acerca de la excitación de las fibras que los neurofisiólogos relacionan con el dolor. Sabemos mucho más del dolor que de la excitación de esas fibras, y lo sabemos desde mucho tiempo antes. Los argumentos epistemológicos contra la identificación de sustancias o procesos concluyen, a partir de un discurso como éste, que dos cosas (sustancias, procesos, eventos) con propiedades distintas en el sentido acabado de exponer no pueden identificarse. En todos los casos se utiliza la Ley de Leibniz en su versión de la indiscernibilidad de los idénticos para rechazar la identificación: si dos cosas tienen propiedades distintas (del tipo que sea, en este caso epistémico) no pueden ser idénticas.

3.1.2. *El argumento de la localización espacial*

Los procesos neuronales ocurren *en* el cerebro. La moderna neurofisiología puede ya localizar con precisión muchos de ellos, y es razonable pensar que, con tiempo suficiente, será capaz de localizarlos todos. Un determinado pensamiento, un deseo, una inquietud, una esperanza, todos estos eventos disparan la actividad en una zona del cerebro, esta correlación puede comprobarse pero la correlación no es identidad. De que dos eventos ocurran siempre a la vez no se sigue que sean idénticos. La subida de la temperatura y el movimiento del mercurio en el termómetro están correlacionados pero son eventos distintos, como lo son las fases de la luna y los movimientos de las mareas. Sin tomar partido por una u otra de las teorías de la relación entre el cuerpo y la mente, es evidente que los procesos cerebrales

ocurren en un sitio concreto, pero ¿dónde exactamente ocurre mi deseo actual de tomar una cerveza fresca con unas sardinas? Un teórico de la identidad diría: exactamente en el mismo sitio donde ocurra la actividad mental que *es* ese deseo. Del mismo modo que Victoria y mi hija de seis años no pueden estar en lugares diferentes porque son una y la misma persona, de acuerdo con la TIP no hay dos eventos ocurriendo, mi deseo de algo y la actividad de mis neuronas, sino uno y el mismo. Pero démonos cuenta de que esta identificación de los lugares de ocurrencia de pensamientos, deseos, esperanzas, etc. y los procesos cerebrales correspondientes no es algo que pueda probarse a partir de un argumento, sino algo que actúa como premisa para la identificación. Si son la misma cosa eventos mentales y eventos cerebrales, coincidirá su localización espacial. En caso contrario, no coincidirá. El argumento contra la TIP utiliza la intuición pre-teórica de que los pensamientos no ocupan lugar, para atribuir propiedades distintas (en este caso espaciales) a los eventos mentales y a los eventos físicos o cerebrales. La estructura del argumento es similar a la de los argumentos epistémicos:

- (i) la activación de mis neuronas se produce en el lugar e_0
- (ii) mi deseo de tomar una cerveza con sardinas no se produce en el lugar e_0 (porque no se produce en ningún sitio)
- (iii) luego, mi deseo y la activación de mis neuronas no pueden ser lo mismo (por la Ley de Leibniz, en su versión de la indiscernibilidad de los idénticos).

3.1.3. *El argumento de las propiedades fenoménicas*

Las propiedades fenoménicas de los objetos, los *qualia*, son las formas en las que los percibimos, la forma en que percibo el amarillo de esa cerveza, el peculiar olor de las sardinas asadas o el sabor de las quisquillas, lo que siento ahora cuando me duele la cabeza, cómo siento una punzada en el estómago, etc. La explicación de los *qualia* es uno de los retos más difíciles para una teoría materialista de la mente-cerebro. Ya hemos visto en el ejemplo de Jackson un argumento que utilizaba los *qualia* contra la TIP. Ahora veremos otro que usa la Ley de Leibniz y las propiedades fenoménicas contra ella. Podemos hablar de un tono de amarillo deslumbrante, de un olor persistente o evocador, de un sabor inclasificable, de un dolor punzante, de un picor desesperante, etc. No tiene sentido, sin embargo, hablar de una excitación punzante de un determinado grupo de neuronas, o de una excitación deslumbrante, persistente, punzante o desesperante. De este mo-

do, sigue el argumento, hay propiedades de los eventos mentales que no son propiedades de los eventos físicos. Si mi dolor es punzante pero no puedo calificar de tal a la excitación de mis fibras C en este momento, entonces mi dolor no es la excitación de mis fibras C. De nuevo tenemos aquí un argumento de estructura similar a los anteriores, que usa la indiscernibilidad de los idénticos contra la TIP.

Las propiedades fenoménicas o *qualia* provocan también otra clase de argumentos de tipo epistémico contra la identidad psicofísica. Estos argumentos tienen que ver con los distintos criterios de aplicación de las expresiones involucradas ("dolor" o "excitación de las fibras C"), en la versión más semántica, o con las distintas formas de conocimiento o de acceso a los procesos mentales y cerebrales, en la versión más propiamente epistémica. Los caminos por los que yo conozco que tengo un dolor son diferentes de aquéllos por los que descubro que mis fibras C están excitadas. Para la aplicación de los términos fenoménicos, o para el conocimiento de la ocurrencia de los eventos mentales, es esencial la distinción entre la perspectiva de la primera persona y la perspectiva de la tercera persona. A mi dolor yo tengo un acceso inmediato, mientras que los demás tienen que inferirlo de mi comportamiento. Es un evento esencialmente privado. La excitación de mis fibras C es, en principio, un evento público: cualquiera que esté en la situación adecuada (contemplando una operación quirúrgica en mi cerebro, por ejemplo) puede constatarlo y para el observador la distinción entre la perspectiva de la primera persona y la perspectiva de la tercera persona es irrelevante. Si quiero ver la excitación de mis fibras C tendré que hacerlo con un método similar al que tendría que usar cualquier otro observador distinto de mí (si me mantuviera despierta durante la operación de mi cerebro, podría observar el proceso con un espejo etc.). Las distintas formas de acceso o los distintos criterios de aplicación de los términos proporcionan propiedades epistémicas y semánticas distintas para el argumento contra la identificación a través, de nuevo, de la Ley de Leibniz.

El argumento de las distintas formas de acceso no se detiene en la constatación de que hay distintas vías para conocer un determinado fenómeno, sino que pretende tener un alcance ontológico. Las distintas formas de acceso, las distintas vías de conocimiento, implican, según los proponentes de este argumento, distintas formas de cognoscibilidad, que indican a su vez que estamos ante procesos distintos. No es sólo que de hecho conocamos las propiedades fenoménicas y las físicas por distintos canales, sino que ambos tipos de procesos, eventos o propiedades son esencialmente

distintos en cuanto a su cognoscibilidad. La tesis que subyace es que las diferencias en cuanto a cognoscibilidad son un síntoma de diferencias ontológicas. Un argumento así ha sido desarrollado por Nagel¹⁷.

3.2. *El argumento de la realizabilidad variable*

El argumento de la realizabilidad variable, junto con la interpretación correcta de los *qualia* y el problema de la conciencia, es otro de los retos serios para la TIP y ha sido defendido, entre otros, por Putnam¹⁸. Dice, a grandes rasgos, lo siguiente; si la TIP fuera correcta, entonces cada tipo de evento mental se identificaría con un determinado tipo de evento neuronal. Esto es, si el dolor *es* la excitación de las fibras C, ningún organismo que no estuviera en posesión de este tipo de fibras podría sentir dolor. La identidad, sigue el argumento, es una relación de uno a uno (una relación un poco peculiar, bien es verdad, la relación que tiene cada cosa consigo misma), mientras que el argumento de la realizabilidad variable parte de la intuición de que la relación que ocurre en este caso es más bien de uno a muchos.

Los eventos mentales son esencialmente eventos privados que son cognoscibles mediante un procedimiento introspectivo. Parece razonable pensar que si alguien (humano o no) sabe que siente dolor, su conocimiento no puede ponerse en duda. Pero el procedimiento introspectivo nada dice acerca del soporte físico del evento mental: podría ser la activación de las fibras C o cualquier otro proceso físico. ¿Por qué descartar *a priori* que seres con una constitución física distinta de la nuestra puedan tener propiedades mentales idénticas a las que nosotros tenemos? El argumento, de ser correcto, no sólo se aplicaría en el caso de diversas especies con distintas estructuras, sino que vale también dentro de la propia especie humana. Dada la plasticidad de nuestro cerebro, es posible que, de un momento al siguiente, el mismo tipo de eventos mentales se materialicen en tipos distintos de eventos físicos. El argumento de la realizabilidad variable tiene la siguiente estructura:

- (i) mi sentir dolor en el tiempo t_1 = la excitación de mis fibras α (distinto de β) en ese tiempo
- (ii) mi sentir dolor en el tiempo t_2 = la excitación de mis fibras β (distinto de α) en ese tiempo
- (iii) luego, mi sentir dolor en t_1 no puede ser mi sentir dolor en t_2 .

Dos consecuencias pueden extraerse de aquí. La primera, que si el dolor en un caso es la excitación de un tipo de células y en otro es la excitación de un tipo de células distinto, entonces no podemos hablar en ambos casos de "dolor" (por la indiscernibilidad de los idénticos). Estaríamos ante dos procesos mentales nombrados con la misma palabra, que sería así ambigua. La segunda, que si el dolor es la excitación de las fibras α y también la excitación de las fibras β , por la transitividad de la identidad, la excitación de α y β son el mismo proceso físico, lo que está descartado por la hipótesis. Quienes aceptan este argumento, concluyen que los procesos mentales-tipo, como el dolor, no pueden identificarse con ningún tipo de proceso o evento cerebral.

Desde el argumento de la realizabilidad variable se ha desembocado históricamente tanto en el dualismo (el dolor no se identifica con ningún tipo particular de estado cerebral), como en el funcionalismo (el dolor es cualquier cosa que cumpla los papeles funcionales adecuados)¹⁹.

3.3. El argumento lógico-semántico de Kripke²⁰

El argumento de Kripke se basa en su concepción semántica general. La tesis básica (2.2) de toda teoría de la identidad decía que las identificaciones de estados mentales con estados neurológicos son identificaciones empíricas. En la década de los sesenta, el momento de mayor esplendor de la TIP, se consideraban como identificaciones empíricas, esto es, contingentes y *a posteriori*, las identificaciones siguientes:

(3.3.1) el agua es H_2O

(3.3.2) el oro es el elemento de número atómico 79, y

(3.3.3) el calor es el movimiento de las moléculas,

que son, en la concepción kripkeana, oraciones de identidad entre designadores rígidos. Usando un argumento de Ruth Barcan, Kripke llega a la conclusión de que todas las identidades entre designadores rígidos son, si verdaderas, necesariamente verdaderas. Luego, si

(3.3.4) el dolor es la excitación de las fibras C

fuera de este tipo, la identificación del dolor con la excitación de esas fibras sería una identificación necesaria (como lo son en el paradigma kripkeano las tres anteriores (3.3.1), (3.3.2) y (3.3.3)). Kripke, sin embargo, rechaza esta conclusión. Su posición es que los enunciados (3.3.1), (3.3.2) y

(3.3.3) expresan verdades necesarias cognoscibles *a posteriori*, mientras que no hay nada necesario en la (aparente) identificación (3.3.4). La razón tiene que ver con la peculiar inmediatez de la que gozan las propiedades mentales.

Es posible imaginar una situación en la que nos encontremos rodeados de un líquido inodoro, incoloro e insípido, que quite la sed y que fluya por ríos y mares y que, no obstante, no sea H₂O. En este caso, Kripke no tiene reparos en afirmar que, independientemente de lo que creamos o sepamos, de lo que nos parezca o estemos inclinados a decir, independientemente de todas las propiedades fenoménicas que este líquido posea, este líquido no es agua, porque el agua es H₂O. Este argumento, sin embargo, está vedado para los procesos mentales, porque en el caso de los procesos mentales sus propiedades fenoménicas son esenciales al evento en cuestión, aquí no hay posibilidad de distinguir entre propiedades fenoménicas y propiedades subyacentes. Si nos parece dolor, *es* dolor, si vemos un determinado tono de verde, *vemos* un determinado tono de verde. Aquí no existe el recurso de pasar de las formas en las que las cosas se nos aparecen a lo que realmente son. De esta argumentación deduce Kripke la falsedad de las teorías de la identidad mente-cerebro y abre así las puertas al dualismo. Nótese que el argumento de Kripke descansa también en la caracterización de la TIP como teoría de la identidad, de otro modo el argumento de Ruth Barcan no sería aquí aplicable.

Los defensores de la TIP han desarrollado recursos y argumentos contra estos aparentes contraejemplos, entre los cuales el más inmediato es negar que los términos singulares, como "mi dolor" o "la excitación de las fibras C", que aparecen en los enunciados de identidad de la TIP sean designadores rígidos. No veremos aquí estas estrategias para frenar las conclusiones contra la TIP, porque para los efectos de nuestra argumentación estas son irrelevantes.

4. El operador de identidad²¹

Entre los muchos supuestos filosóficos que subyacen a las discusiones anteriores, hay dos especialmente cruciales. El primero es que la TIP es una teoría de la identidad, esto es, que sus afirmaciones básicas son enunciados de identidad de la forma: "el evento mental α es el mismo evento que el evento neuronal β ". El segundo es que el operador de identidad que aparece en estos supuestos enunciados de identidad es el relator que usamos habitualmente ("=") en el Cálculo de Predicados de Primer Orden y que tiene

la categoría lógica de un relator diádico cuyos argumentos son términos singulares. Lo que la caracterización de la identidad como un relator diádico de primer orden significa es que es una expresión que, para forjar una oración bien formada, necesita dos argumentos que sean términos singulares. Un relator de este tipo es, por ejemplo, "...lucha contra..." que necesita dos argumentos para expresar una proposición como "El Imperio lucha contra la Rebelión". Además de las expresiones deícticas y los pronombres, que no suelen ser objeto de tratamiento específico en el Cálculo de Predicados estándar, dos son los tipos de términos singulares que pueden encajar en los lugares de argumento de los predicados y las relaciones de primer orden: los nombres propios y las descripciones. El relator "...encontrarse con..." sirve para formar oraciones del tipo: (i) "Lúthien se encontró con Beren", (ii) "Lúthien se encontró con un hombre", o (iii) "Lúthien se encontró con el hombre de su vida", que tienen respectivamente, la estructura:

- (i) nombre + relator + nombre,
- (ii) nombre + relator + descripción indefinida, y
- (iii) nombre + relator + descripción definida.

Con el signo de identidad convencional la opción (ii) no es posible: "Eärendil es un hombre" no es una oración de identidad, sino una predicación en la que "es" no es ningún operador separable sino que forma parte del predicado "es un hombre". Atendiendo a esto, hay tres posibles estructuras que las oraciones de identidad del Cálculo de Predicados pueden poseer:

- (a) nombre = nombre,
- (b) nombre = descripción definida, y
- (c) descripción definida = descripción definida.

Veamos ejemplos de estos tres tipos de estructuras.

- (a.i) nombre = nombre
Thingol es Elwë
- (b.i) nombre = descripción definida
Thingol es el padre de Lúthien
- (c.i) descripción definida = descripción definida
El rey de Doriath es el padre de Lúthien

En lo que ahora estamos interesados es en la forma de los enunciados de identidad o, como preferimos decirlo, en los componentes y la estructura de las proposiciones expresadas por ellos. El mejor camino para conseguirlo consiste en analizar qué es exactamente lo que los hablantes quieren transmitir cuando utilizan enunciados como (a.i), (b.i) y (c.i).

Si un hablante utiliza una oración como (a.i) puede querer decir algo tautológico, que no aumente nuestro conocimiento, pero seguramente los enunciados de identidad entre nombres propios tienen alguna utilidad más interesante que la de repetir lo que todo el mundo ya sabe. Estamos hablando de enunciados de identidad entre *nombres propios genuinos*, no entre expresiones sintácticamente simples que en realidad funcionen como descripciones. En este caso los términos no pueden ser del tipo de los fregeanos *Morgenstern* y *Abendstern*. La única opción compatible con un Principio de Caridad que suponga que el hablante está transmitiendo algo de interés consiste en interpretar (a.i) como una afirmación metateórica en la que uno de sus dos nombres (o a los dos) aparezca mencionado:

- (a.ii) Thingol también es conocido como "Elwë", o
- (a.iii) "Thingol" y "Elwë" refieren al mismo individuo.

Si esta interpretación es aceptable, ¿en qué sentido puede decirse que (a.i) realiza una afirmación de identidad? ¿De qué objeto u objetos decimos que son idénticos? No de los nombres, desde luego. Más bien lo que afirmamos con (a.i) es que un determinado individuo tiene dos nombres o, alternativamente, que esos dos nombres son co-referenciales. En este caso la supuesta afirmación de identidad se ha desvanecido.

Veamos qué ocurre con (b.i). Si tuviéramos que formalizar esta oración en el Cálculo de Predicados, la representaríamos como " $a = \text{ix Pxb}$ ". ¿Qué diferencia habría entre esta formalización y la más simple "Pab"? Hay una diferencia, pero no es demasiado grande. Lo que añade la combinación del signo de identidad más el descriptor a la formulación más simple no es más que la afirmación de que el predicado "P...b" sólo puede predicarse de un único individuo, a saber, del individuo al que llamamos "a". Lo que se dice es que sólo Thingol es el padre de Lúthien. Estas consideraciones nos dan la pauta para la interpretación correcta de (b.i), a saber, que con esa oración se afirman las siguientes dos cosas:

- (b.ii) Lúthien tiene un padre, a saber, Thingol y
- (b.iii) Lúthien tiene únicamente un padre.

Si traducimos al Cálculo de Predicados (b.ii) obtendremos una fórmula como "Pab", si traducimos (b.iii) lo que obtendremos será una fórmula como "Pab \wedge $\neg \exists xy$ (Pxb \wedge Pyb)", adoptando la interpretación exclusiva de las variables propuesta por Wittgenstein. ¿Dónde aparece aquí la afirmación de identidad, en la simple relación de (b.ii), o en la relación más la afirmación de unicidad de (b.iii)? La respuesta con premio es: "en ninguna de las dos". No hay afirmaciones de identidad en las oraciones en las que aparece un nombre y una descripción. No afirmamos la identidad de nada si decimos que Aegnor es el único gemelo de Angrod. Decimos que Aegnor tiene un hermano gemelo y que Aegnor no tiene más gemelos aparte de Angrod. También en este caso la supuesta relación de identidad se nos ha ido por el desagüe del análisis lógico.

El análisis del ejemplo (c.i) es algo más complejo. Las descripciones definidas expresan, por un lado, una predicación: decimos de algo o de alguien que tiene una determinada propiedad. En el caso de nuestro ejemplo, decimos que alguien es el padre de Lúthien y decimos también que alguien es el Rey de Doriath. Con las descripciones definidas decimos también algo más, a saber, que esa propiedad se ejemplifica en un único individuo, esto es, que Lúthien sólo tiene un padre y que sólo hay un rey en Doriath. Hasta aquí lo que las descripciones definidas expresan. Pero, a diferencia de lo que ocurre en (a.i) y (b.i), el ejemplo (c.i) es una afirmación genuina de identidad, aunque hasta ahora, al explicar la contribución semántica de las descripciones no hemos hecho uso de esta noción. Cada descripción definida expresa predicación y unicidad pero, cuando se unen dos descripciones definidas en oraciones como (c.i), se afirma además que el individuo que únicamente ejemplifica la primera propiedad es el mismo que el individuo que únicamente ejemplifica la segunda. Detengámonos un momento en este punto y veamos cuáles son los argumentos que apoyan la interpretación que hace de (c.i) una afirmación de identidad. Para ello haremos uso de dos importantísimas intuiciones del Wittgenstein del *Tractatus logico-philosophicus*: el uso exclusivo de las variables (una variable para cada objeto y no dos objetos con la misma variable) y la distinción entre mostrar y decir.

El ejemplo (c.i) puede reformularse de la siguiente manera:

(c.ii) x es el padre de Lúthien y x es Rey de Doriath,

donde la repetición de la variable x muestra la identidad. Vemos que predicamos las dos propiedades del mismo individuo porque vemos que apa-

rece dos veces la misma variable. Pero si queremos *decir* explícitamente que sólo hay un individuo involucrado, entonces necesitamos un operador nuevo, un operador de identidad que afirme que las dos descripciones se instancian únicamente en *el mismo* individuo. Sin embargo, ese operador, el operador de identidad propiamente dicho, no es un operador de primer orden, como el que se utiliza habitualmente en el Cálculo de Predicados, sino un operador de orden superior, esto es, un operador cuyos argumentos son predicados, un operador del tipo de los cuantificadores. El error de considerar al operador de identidad como un operador de primer orden es el responsable de las paradojas y dificultades formales y filosóficas que se plantean en la teoría de la identidad. Volvamos al ejemplo. En (c.i) no tenemos dos objetos del que decimos que es el mismo, lo que tenemos es un objeto al que le adscribimos dos propiedades. Lo que en (c.i) se dice de un individuo podría decirse también de dos, el relator

(d.i) ... es el padre de Lúthien y ... es rey de Doriath

tiene dos lugares de argumento, que podrían rellenarse mediante dos argumentos distintos. Por ejemplo,

(d.ii) Thingol es el padre de Lúthien y Eärendil es rey de Doriath,

es una oración falsa, pero bien formada.

Pero el relator diádico (d.i) puede utilizarse para decir esas dos cosas de un único individuo, y en este caso la oración puede construirse a partir de sus elementos básicos de dos maneras distintas. Bien mediante un relator diádico cuyos dos lugares de argumento se saturan repitiendo el mismo argumento, o bien mediante la construcción de un predicado conjuntivo, a partir del predicado diádico original, usando un operador que clausure los dos lugares y abra otro. Unas pocas palabras de aclaración no estarán de más aquí. Los ingredientes básico de (c.i) son dos predicados monádicos, "... es padre de Lúthien" (lo simbolizaremos por "P") y "... es rey de Doriath" (lo simbolizaremos por "D"), un nombre "Thingol" (al que simbolizaremos por "a") y el signo de conjunción. La forma lógica de (c.i) podría ser una de las dos siguientes, ambas extensionalmente equivalentes, pero que indican vías de construcción distintas y una es más informativa que la otra. Podríamos construir dos oraciones elementales, cada una con su predicado monádico y su término singular, "Pa" y "Da", y a continuación unir las mediante el signo de conjunción, "Pa \wedge Da", o unir los dos predicados "P... \wedge D..." mediante la conjunción, convertir el relator conjuntivo

diádico resultante en un predicado monádico, " $(P \wedge D) \dots$ ", y saturar su lugar de argumento con el término que teníamos, " $(P \wedge D)a$ ". Esta es la idea, pero el paso de " $P \dots \wedge D \dots$ " a " $(P \wedge D) \dots$ " no puede hacerse así. La conversión del relator diádico en un predicado monádico exige la aparición de un operador nuevo, capaz de clausurar lugares de argumento y convertir predicados en otros más complejos.

Veamos todo esto con algunos ejemplos. Un ejemplo de " $P \dots \wedge D \dots$ " sería (d.i), y un ejemplo de " $(P \wedge D) \dots$ " sería:

(e.i) ... (es-el padre-de Lúthien-y-rey-de-Doriath).

La diferencia entre (d.i) y (e.i) consiste en que el primero es un relator diádico mientras que el segundo es monádico. Con (d.i) podemos formar oraciones como

(d.ii) Thingol es padre de Lúthien y Eärendil es rey de Doriath,

o como

(d.iii) Thingol es el padre de Lúthien y Thingol es rey de Doriath,

mientras que con (e.i) no tenemos esa posibilidad. Con un relator monádico sólo podemos tener un argumento y, por tanto, sólo oraciones como

(e.ii) Thingol es el padre de Lúthien y rey de Doriath.

En contextos extensionales (d.iii) y (e.ii) no presentan diferencias lógicas relevantes (aunque la proposición expresada por ambas no sería la misma) y se seguirían del mismo conjunto de proposiciones e implicarían el mismo conjunto de proposiciones. Sin embargo, (e.ii) tiene un operador que (d.iii) no tiene, un operador que permite convertir el relator diádico del que partíamos, (d.i), en un operador monádico, (e.i). Este operador nuevo, un operador que convierte predicados n -ádicos en predicados $n-1$ -ádicos, puede representarse de muchas maneras. Nosotros elegiremos el operador de identidad de Williams, simbolizado por " \equiv " en honor del Frege de *Begriffsschrift*:

(h) $\equiv u$ (u es el padre de Lúthien $\wedge u$ es rey de Doriath)...

esto es " $\equiv u (Pu \wedge Du) \dots$ ".

La característica relevante de este operador para el argumento general que estamos desarrollando es su estatus de operador de segundo orden, esto es, de operador que exige como argumentos predicados en lugar de términos.

Ha llegado el momento de retomar los diversos hilos argumentales que han ido apareciendo a lo largo de estas páginas y aventurar una conclusión que muestre cómo aplicar el análisis lógico de la noción de identidad a las disputas en filosofía de la mente. Realizaremos esta tarea final en dos pasos. El primero consiste en determinar cuáles son, desde un punto de vista lógico, los ingredientes básicos exigidos para expresar la tesis de la identidad psiconeural. El segundo supone dilucidar qué consecuencias se siguen de la interpretación de la identidad como operador de segundo orden. Finalmente, y esto sólo como un apunte de trabajo, habría que analizar qué es lo que realmente se quiere decir con la tesis de la identidad psiconeural y si esta puede todavía enunciarse de alguna manera. Todo esto se tratará en las secciones que siguen.

5. La identidad psiconeural y el relator de identidad

La teoría de la identidad psiconeural tal como se discute en la actualidad, no suele formularse en términos de la identidad de sustancias o de objetos sino de tipos de eventos o procesos. Cuando queremos referirnos a eventos, no solemos utilizar nombres propios, sino que los identificamos usando otras expresiones complejas. Por esta razón las disquisiciones que hemos llevado a cabo en la sección anterior no se aplican automáticamente a la teoría de la identidad psiconeural. Son eventos

V.1. Mi comentar en este momento que la teoría de la identidad psiconeural es falsa, o

V.2. Mi sentir en este instante un cierto cosquilleo en el estómago.

Puedo referirme a estos eventos de la forma en la que lo hemos hecho en V.1 y V.2 o de otras formas alternativas, como

V.3. Lo que sorprendió tanto a Alberto, o

V.4. Lo que me hizo sentirme tan incómoda.

Respecto del primer bloque no tiene sentido preguntarse cuál es el evento al que me estoy refiriendo porque éste queda transparentemente representado en la manera de designarlo, a saber, mi comentar en este momento

que la TIP es falsa o mi sentir un cosquilleo en el estómago, y desde esta manera de designarlo podemos reconstruir la oración que exprese claramente el evento al que me estoy refiriendo:

V.5. Yo comento en este momento que la teoría de la identidad psico-neural es falsa, o

V.6. Siento en este instante un cierto cosquilleo en el estómago.

Del segundo bloque no pueden reconstruirse, sin más información, las oraciones que expresan de manera más clara los eventos a los que nos estamos refiriendo, porque V.3 y V.4 son generalizaciones de eventos, tienen la forma de descripciones definidas y son compatibles con eventos distintos. Lo que sorprendió tanto a Alberto pudo ser mi comentar eso acerca de la teoría de la identidad o la aparición súbita en la sala de Conferencias sala del Presidente del Gobierno. Lo que me hizo sentirme tan incómoda pudo ser mi sentir ese cosquilleo en el estómago o mi darme cuenta, en mitad de mi alocución, de que había perdido una transparencia. Por eso, para este segundo bloque tiene sentido preguntar, además, a qué evento nos estamos refiriendo. Por esa razón, Williams denomina a las expresiones de este segundo tipo *namely-riders*. Son expresiones que pueden completarse con una cláusula "a saber":

V.7. Lo que sorprendió tanto a Alberto, a saber, mi comentar que la teoría de la identidad psico-neural es falsa, o

V.8. Lo que me hizo sentirme tan incómoda, a saber, mi sensación de cosquilleo en el estómago.

Williams denomina a las expresiones del primer tipo (V.1 y V.2) *designaciones directas* de eventos y a las del segundo tipo (V.3 y V.4) *designaciones indirectas*. Las designaciones indirectas son descripciones definidas mientras que las designaciones directas harían, para el caso de los eventos, las veces de nombres propios. Naturalmente las descripciones directas de eventos no funcionan exactamente igual que los nombres de objetos, pero para las necesidades de esta argumentación la identificación es suficiente.

De la teoría de la identidad expuesta en la sección anterior se sigue que no pueden darse afirmaciones genuinas de identidad entre nombres propios. Al ser el operador de identidad un operador de orden superior, sus argumentos no pueden ser, desde un punto de vista lógico, más que predicados de algún tipo. Afirmamos identidad cuando decimos que la misma persona que fue el padre de Lúthien fue el rey de Doriath, a saber, Thingol.

Las descripciones definidas son también *namely-riders*, expresiones que admiten una cláusula "a saber", y en este sentido son generalizaciones existenciales. "La misma persona que fue el padre de Lúthien fue el rey de Doriath" es una generalización existencial de "Thingol fue el padre de Lúthien y Thingol fue el rey de Doriath". Afirmamos identidad porque decimos que las dos descripciones tienen la misma cláusula "a saber". En el discurso de eventos la teoría es análoga. Afirmamos identidad cuando decimos que dos designaciones indirectas de eventos poseen la misma cláusula "a saber". Así,

V.9. Lo que sorprendió tanto a Alberto es lo que fue rechazado energí-
camente por Juan,

es un enunciado de identidad genuino entre eventos porque a continuación podemos decir "a saber, mi comentar que la TIP es falsa". La forma lógica de V.9, siguiendo el esquema de la sección anterior, sería

V.10. $(\exists\beta) (\equiv\delta)(\delta \text{ sorprendió mucho a Alberto} \wedge \delta \text{ fue energicamente rechazado por Juan})\beta$,

que es una generalización existencial y una afirmación de identidad realizadas sobre

V.11. Mi comentar que la TIP es falsa sorprendió mucho a Alberto \wedge mi comentar que la TIP es falsa fue rechazado energí-
camente por Juan.

Sin embargo, las afirmaciones estándar de la TIP no son como V.9 sino como

(2.2.1) Mi dolor actual es la estimulación de mis fibras C en este momento.

¿De qué conjunción de eventos es (2.2.1) una generalización existencial? La respuesta más sensata aquí es: de ninguna. Naturalmente, siempre se pueden forzar los ejemplos para que encajen en la teoría y decir que (2.2.1) tiene la siguiente forma lógica:

V.12. $(\exists\beta) (\equiv\epsilon) (\epsilon \text{ es mi dolor actual} \wedge \epsilon \text{ es la estimulación de mis fibras C en este momento})\beta$.

Pero ¿qué tipo de evento o proceso sería aquí β ? Podríamos adaptar la teoría de la identidad de la sección anterior a (2.2.1) al precio de convertir "...es mi dolor actual" y "...es mi actual estimulación de las fibras C" en predicados. Sin embargo, desde un punto de vista lógico-semántico esta opción es difícilmente defendible, ya que "mi dolor en este momento" se usa como una expresión referencial y, por tanto, como un término singular y no como un predicado. Si interpretásemos estas expresiones como predicados estaríamos ofreciendo una teoría del Doble Aspecto y abrazando un Monismo Neutral.

La mayor parte de las afirmaciones de la TIP son como (2.2.1), que afirma identidad uniendo dos designaciones directas de eventos. Si para éstas no es posible la interpretación V.12, ¿cómo las interpretaremos? Tampoco podemos analizarlas siguiendo la pauta

(a.i) nombre = nombre

ya que en este caso lo que decimos es que dos expresiones son sinónimas, esto es, que dos nombres propios, cuya única misión semántica es referir, refieren al mismo referente, mientras que "mi dolor" y "la estimulación de las fibras C" no son expresiones sinónimas. Aquí podemos recurrir a la distinción fregeana entre sentido y referencia, que es la interpretación de la teoría dada por Feigl²², pero en este caso no podemos decir que la teoría lleve a cabo ninguna afirmación de identidad, sino a lo sumo de co-referencialidad. Dos designaciones directas de eventos, al igual que dos nombres propios, no son los argumentos apropiados para un operador de identidad. De todos modos, interpretar "mi dolor actual" y "la estimulación de mis fibras C" como nombres fregeanos, esto es, como expresiones dotadas de sentido y referencia, las convierte en expresiones de propiedades, en predicados de algún tipo, y esta interpretación no es fácilmente conciliable con su uso como designaciones directas, que es esencialmente un uso referencial.

El argumento que estamos aquí desarrollando tiene ecos del argumento kripkeano contra la necesidad de las identidades psiconeurales. El argumento de Kripke es metafísico, las propiedades fenoménicas no tienen un aspecto sentido o percibido junto con una esencia, sino que para ellas vale el dictum berkeleyano *esse est percipii*. Nuestro argumento es lógico-semántico, las expresiones que aparecen en los enunciados básicos de la TIP son designaciones directas de eventos, esto es, expresiones cuyo papel semántico no es descriptivo sino referencial. Si estas expresiones no pueden

predicarse de eventos sino que simplemente los señalan, entonces no hay posibilidad de formular la teoría como una teoría de la identidad usando un operador genuino de identidad como " \equiv ", por contraste con el conocido " $=$ ".

6. Conclusión

¿Qué forma lógica tienen los enunciados básicos de la TIP, dado que no son enunciados de identidad genuinos? Esta pregunta requeriría un análisis lógico de la teoría más profundo que el meramente negativo que hemos llevado a cabo en estas páginas. Esa será tarea de los lógicos. ¿Qué quieren decir los proponentes de la TIP cuando hacen sus afirmaciones de pseudo-identidad? Esta tarea de dilucidación recaerá sobre sus espaldas.

Pero no todas las conclusiones de este artículo son negativas. He aquí la conclusión positiva más evidente: si la TIP no es, desde un punto de vista lógico, una teoría de la identidad, los argumentos esgrimidos contra ella que sacan partido de la Ley de Leibniz y del resto de los axiomas que conforman una teoría de este tipo no consiguen tocarla. La TIP, de este modo, no queda refutada por la batería de argumentos expuesta en la sección 3 y por tanto, en teoría, sus afirmaciones básicas, formuladas correctamente, pueden seguir defendiéndose.

Hasta el momento, por tanto, no se ha mostrado que la TIP sea falsa, sólo que hay que descubrir aún qué es lo que la teoría afirma realmente.

Notas

† Una versión anterior de este trabajo fue presentada en un Curso de la Universidad del Mediterráneo, en Motril, celebrado del 7 al 11 de Septiembre de 1998, organizado por los profesores de la Universidad de Granada José Manuel Martín Morillas y Alberto Morales. Quiero agradecerles a ambos su invitación a participar en el curso, la confianza que siempre han depositado en mí y su continuo apoyo. Quiero expresar así mismo mi agradecimiento a mis colegas Juan José Acero y Esther Romero por sus valiosos comentarios y sugerencias al texto, como también las precisiones y modificaciones aconsejadas por los informantes anónimos de la revista *Theoria*.

¹ También se han propuesto en la historia de la filosofía monismos de tipo espiritualista o mentalista, aunque no los contemplaremos en este trabajo.

² Por esta razón al dualismo, como se definirá aquí, se le denomina también "Dualismo de Sustancias", en contraposición a una concepción más débil, el Dualismo de Propiedades, que aceptando la existencia de un único tipo de sustancia, considera no obstante que las propiedades mentales son diferentes de las propiedades físicas y no pueden reducirse a ellas.

- 3 Aunque también hay propuestas dualistas en las que se niega la interacción mente-cerebro y otras, como las que se engloban bajo el rótulo de "Epifenomenalismo", en las que sólo se admite la influencia de lo físico sobre lo mental, pero no al contrario.
- 4 K. Popper y J. Eccles: 1985, *El Yo y su Cerebro*, Barcelona, Labor. Edición original: *The Self and its Brain*, New York, Springer-Verlag, 1977; J. Foster: 1991, *The immaterial self. A defence of the cartesian dualist conception of the mind*. Londres y Nueva York, Routledge; R. Swinburne: 1986, *The evolution of the soul*, Oxford, Clarendon Press.
- 5 Para una análisis más detallado de todas estas posiciones vid., por ejemplo, Samuel Guttenplan (ed.): 1995, *A Companion to the Philosophy of Mind*, Blackwell, Oxford; Georges Rey: 1997, *Contemporary Philosophy of Mind*, Blackwell, Oxford; y E. Rabossi: 1995, 'La tesis de la identidad mente-cuerpo', in F. Broncano (ed.): *La mente humana*, Madrid, Ed. Trotta.
- 6 Esta última es la interpretación que hace Feldman de la postura de Leibniz. Vid. F. Feldman: 1970, 'Leibniz and Leibniz' Law', *Philosophical Review* LXXIX/2.
- 7 U.T. Place: 1956, 'Is consciousness a brain process?', J.J. Smart: 1959, 'Sensations and brain process', ambos incluidos en C.V. Borst (ed.): 1970, *The mind/brain identity theory*, The Macmillan Press Ltd., y H. Feigl: 1980, 'Mind-body, not a pseudo problem', in N. Block (ed.): *Readings in the Philosophy of Psychology*, Cambridge, Harvard University Press, pp. 148-155.
- 8 J.B. Watson: 1913, 'Psychology as the Behaviorist Views It', *Psychological Review* 20, 158-177; B.F. Skinner: 1953, *Science and Human Behavior*, New York, Macmillan; G. Ryle: 1949, *The Concept of the Mind*, New York, Barnes and Noble.
- 9 G. Ryle: 1949, *The Concept of the Mind*, New York, Barnes and Noble. Hay traducción castellana: *El Concepto de lo Mental*, Buenos Aires, Paidós.
- 10 Véase D.M. Armstrong: 1970, 'The Nature of Mind', in C.V. Borst (ed.): *The Mind/Brain Identity Theory*, London, The Macmillan Press, y *A Materialist Theory of Mind*, London, Routledge & Kegan Paul, 1968.
- 11 La versión lógico-semántica fue defendida por R. Carnap en *The logical structure of the world*, Berkeley, University of California Press, 1969.
- 12 Frank Jackson: 1986, 'What Mary Didn't Know', *The Journal of Philosophy* LXXXIII/5, 291-5.
- 13 Para una discusión más detallada de las posibles actitudes ante el experimento de Jackson, véase Rovert van Gulick: 1997, 'Understanding the Phenomenal Mind: Are we all just Armadillos? Part I: Phenomenal Knowledge and Explanatory Gaps.', in Ned Block, Owen Flanagan y Güven Güzeldere (eds.): *The Nature of Consciousness*, Cambridge, Mass., The MIT Press, pp. 559-566.
- 14 R. Descartes: 'Meditación Segunda', in *Meditaciones Metafísicas con objeciones y respuestas*, Introducción, traducción y notas de Vidal Peña, Madrid, Ediciones Alfaguara, 1977.
- 15 En R. Descartes: 'Meditación Sexta', *op. cit.* Para un análisis detallado de los dos pasos de este argumento cartesiano, véase B. Williams: 1996, *Descartes. El proyecto de la investigación pura*, Madrid, Cátedra, especialmente el capítulo IV, 'La Distinción Real'. Agradezco a un anónimo informante de *Theoria* el haberme hecho caer en la cuenta de esta distinción.

- 16 Exposiciones muy claras de los argumentos más comunes a la teoría de la identidad psiconeural se encuentran en Jaegwon Kim: 1996, *Philosophy of Mind*, Westview Press, cap. 3, y en Georges Rey: 1997, *Contemporary Philosophy of Mind*, Blackwell, Oxford, cap. 2.
- 17 T. Nagel: 1974, 'What is it like to be a bat?', *The Philosophical Review* 83, 435-450.
- 18 H. Putnam: 1967, 'Psychological Predicates', in W.H. Capitan y D.D. Merrill (eds.): *Art, mind, and religion*, Pittsburgh, University of Pittsburgh Press, pp. 37-48.
- 19 En Jaegwon Kim: 1996, *Philosophy of Mind*, Westview Press, se presentan los argumentos tradicionales a favor del dualismo (cap. 3) y del funcionalismo (cap. 4) que se derivan de la supuesta refutación de la teoría de la identidad del tipo sobre la base del argumento de la realizabilidad variable. No obstante, en otros sitios, Kim ofrece interpretaciones del argumento de la realizabilidad variable que lo inutilizan como apoyo tanto al dualismo como al funcionalismo. Vid., por ejemplo: 'Multiple realization and the metaphysics of reduction', in J. Kim: *Supervenience and Mind*, Cambridge, Cambridge University Press. Agradezco a un anónimo *referee* de *Theoria* el que me haya llamado la atención sobre este punto.
- 20 Vid. S. Kripke: 1995, 'Identidad y Necesidad', in L. Valdés (ed.): *La Búsqueda del Significado*, Madrid, Tecnos. Versión original en Milton K. Munitz: 1971, *Identity and Individuation*, New York, New York University Press. También en S. Kripke: 1985, *El Nombrar y la Necesidad*, México D.F., U.N.A.M. Versión original: *Naming and Necessity*, Oxford, Basil Blackwell Publisher, 1980.
- 21 La teoría de la identidad que se expondrá en esta sección y en la siguiente se debe básicamente a C. Williams. Vid. C. Williams: 1989, *What is identity?*, Oxford, Clarendon Press, y *Being, Identity, and Truth*, Oxford, Clarendon Press, 1992, especialmente el cap. VII, titulado 'Pains and Brains'.
- 22 Vid. H. Feigl: 1958, 'The "Mental" and the "Physical"', in H. Feigl, M. Scriven y G. Maxwell (eds): *Minnesota Studies in the Philosophy of Science II*, Minneapolis, University of Minnesota Press.

María José Frápolli es Profesora titular de Lógica y Filosofía de la Ciencia en el Departamento de Filosofía de la Universidad de Granada. Ha realizado estancias de investigación en las Universidades de Warwick y Bristol (Reino Unido) y en la Universidad de Miami (USA). Ha publicado artículos en las revistas *Modern Logic*, *History and Philosophy of Logic*, *Agora*, *Revista de Filosofía*, *Teorema*, *Diálogo Filosófico*, etc. Ha editado en colaboración con Juan A. Nicolás el libro *Teorías de la Verdad en el Siglo XX* (Tecnos, 1997) y *Verdad y Experiencia* (Comares, 1999), y es autora junto con Esther Romero de: *Una aproximación a la Filosofía del Lenguaje* (Síntesis, 1998).