La filosofía de la mente ocupa un lugar único entre las cuestiones filosóficas contemporáneas, por cuanto la totalidad de las teorías más célebres e influyentes son falsas. Uno de mis objetivos es intentar rescatar la verdad del abrumador influjo de la falsedad. He procurado llevar a cabo parte de esta tarea en otras obras, especialmente en The Rediscovery of the Mind, *pero el presente es mi único intento de escribir una introducción general que abarque el tópico de la filosofía de la mente en su conjunto.*

DE LA INTRODUCCIÓN, JOHN R. SEARLE

Searle ha escrito un libro introductorio sólido, claro, accesible y fascinante, que explica de manera mucho más convincente que cualquier otro su concepción iconoclasta de que tanto el materialismo como el dualismo son falsos. El autor se embarca en una vigorosa exploración de las grandes cuestiones de la filosofía de la mente, siempre concentrado en las intuiciones más profundas sobre el tema.

NED BLOCK, UNIVERSIDAD DE NUEVA YORK

vitral

www.norma.com

9 789580 492443
Código 22332
ISBN 958-04-9244-1

John R. Searle LA MENTE

GRUPO EDITORIAL norma

vitral

John R. Searle

LA MENTE

Una breve introducción

GRUPO EDITORIAL norma

John R. Searle, la principal autoridad en temas de la mente, ofrece una cautivante introducción a esta, una de las zonas más enigmáticas de la filosofía, a través de una discusión franca y directa, que recorre los conocimientos aceptados al mismo tiempo que propone sorprendentes nuevas ideas sobre la naturaleza de la conciencia y la mente.

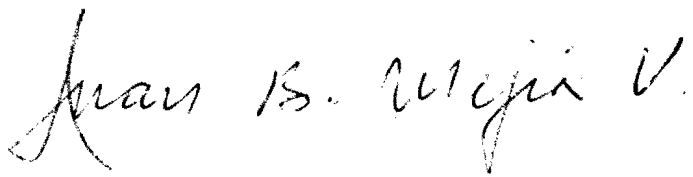
John R. Searle es profesor titular de la cátedra Mills del Departamento de Filosofía de la Universidad de California en Berkeley. Es autor de numerosos libros, entre los que cabe destacar The Rediscovery of the Mind, The Mystery of Consciousness, Mind, Language and Society, Philosophy in the Real World y Consciousness and Language.

John R. Searle

LA MENTE

UNA BREVE INTRODUCCIÓN

Traducción de Horacio Pons



John R. Searle

LA MENTE

UNA BREVE INTRODUCCIÓN

Traducción de Horacio Pons

GRUPO EDITORIAL NORMA

www.norma.com

Bogotá Barcelona Buenos Aires Caracas
Guatemala Lima México Panamá Quito
San José San Juan San Salvador
Santiago de Chile Santo Domingo

Searle, John R.
La mente : una breve introducción / John R. Searle ; traducción Horacio Pons. -- Bogotá : Grupo Editorial Norma, 2006.
382 p. ; 23 cm. -- (Colección Vitral)
Título original. Mind : A Brief Introduction.
ISBN 958-04-9244-1
1. Ciencia cognoscitiva 2. Filosofía de la mente 3. Mente y cuerpo 4. Voluntad (Psicología) I. Pons, Horacio., tr. II. Tít. III. Serie.
128.2 cd 19 ed.
A1076983

CEP-Banco de la República-Biblioteca Luis Ángel Arango

Apartado Aéreo 53550, Bogotá - Colombia
Primera edición, abril de 2006

Impreso por Nomos S.A.
Impreso en Colombia - *Printed in Colombia*

Diseño de cubierta: Camilo Umaña
Ilustración de cubierta: Olga Lucía García
Armada: Blanca Villalba Palacios

CC 22332
ISBN 958-04-9244-1

Este libro se compuso en caracteres Berkeley

CONTENIDO

Para Dagmar

AGRADECIMIENTOS

He presentado la mayor parte del material de este libro en conferencias pronunciadas en Berkeley. Estoy en deuda con mis alumnos por su actitud entusiasta y escéptica a la vez. Dos de ellos, Hua (Linda) Ding y Nadia Taylor, leyeron todo el manuscrito e hicieron útiles comentarios. Por su ayuda en la preparación del texto electrónico, también estoy agradecido con Maria Francisca Reines, Jessica Samuels y Jing Fong Williams Ying. Recibí valiosos consejos filosóficos de Janet Broughton, Josef Moural, Axel Seeman y Marga Vega. Los dos lectores de Oxford University Press, David Chalmers y otro cuya identidad ignoro, plantearon numerosas observaciones de utilidad. Agradezco a mi asistente de investigación, Jennifer Hudin, por su colaboración en todas las etapas del libro, desde la formulación inicial de las ideas hasta la finalización del libro. Y, sobre todo, debo agradecer a mi esposa Dagmar Searle por su consejo y apoyo constantes; el libro está dedicado a ella.

INTRODUCCIÓN

Por qué escribí este libro

En los últimos tiempos se han publicado muchos libros introductorios sobre la filosofía de la mente. Varios de ellos hacen una revisión más o menos amplia de las principales posiciones y argumentos actualmente vigentes en ese campo. Algunos, en verdad, están escritos con gran claridad, rigor, inteligencia y erudición. ¿Cuál es, entonces, mi excusa para añadir un libro más a ese repertorio? Bien, es improbable, desde luego, que un filósofo que haya trabajado con ahínco sobre un tema se sienta completamente satisfecho con los escritos de otro acerca de ese mismo tema; supongo que en ese aspecto soy un filósofo típico. Pero además del deseo habitual de exponer mis desacuerdos, la ambición de escribir una introducción general a la filosofía de la mente se explica por una razón preponderante. Casi todas las obras que he leído aceptan la herencia histórica del mismo grupo de categorías para describir los fenómenos mentales, en especial la conciencia, y con ellas, también un conjunto específico de supuestos sobre las relaciones de la conciencia y otros fenómenos mentales entre sí y con el resto del mundo. Lo que carece de todo cuestionamiento y mantiene la vigencia de la discusión es ese conjunto de categorías, y los supuestos que ellas acarrean como un pesado equipaje. Las diferentes posiciones, por lo tanto, se plantean en el marco de una serie de supuestos erróneos. Como resultado, la filosofía de la mente ocupa un lugar único entre las cuestiones filosóficas contemporáneas, por cuanto la totalidad de las teorías más célebres e influyentes son falsas. Cuando hablo de teorías me refiero sencillamente a todo lo que se

designa con un "ismo". Y pienso en el dualismo –tanto de las propiedades como de las sustancias–, el materialismo, el fisicalismo, el computacionalismo, el funcionalismo, el conductismo, el epifenomenalismo, el cognitivismo, el eliminativismo, el panpsiquismo, la teoría del doble aspecto y el emergentismo, tal como suele concebírselo. Para hacer del tema algo aún más vital, varias de esas teorías, sobre todo el dualismo y el materialismo, tratan de decir algo cierto. Uno de mis muchos objetivos es intentar rescatar la verdad del abrumador influjo de la falsedad. He procurado llevar a cabo parte de esta tarea en otras obras, especialmente en *The Rediscovery of the Mind*[1], pero el presente es mi único intento de escribir una introducción general que abarque el tópico de la filosofía de la mente en su conjunto.

Ahora bien, ¿cuáles son exactamente esos supuestos y por qué son falsos? Todavía no puedo decirlo. No admiten una rápida síntesis sin un trabajo preliminar. La primera mitad de este libro se dedica en gran parte a exponerlos y superarlos. Es difícil resumirlos porque carecemos de un vocabulario neutral para describir los fenómenos mentales. Tengo que comenzar, por ende, apelando a las experiencias de mi lector. Supongamos que usted está sentado a la mesa y piensa en la situación política contemporánea y lo que sucede en Washington, Londres y París. Ahora pone la atención en este libro y lee hasta aquí. En este punto sugiero que, para tener una idea de los supuestos, trate de pellizcarse el brazo izquierdo con la mano derecha. Y suponga que lo hace intencionalmente. Esto es,

1 J. R. Searle, *The Rediscovery of the Mind*, Cambridge (Mass.), MIT Press, 1992 [traducción española: *El redescubrimiento de la mente*, Barcelona, Crítica, 1996].

supondremos que su intención causa el movimiento que lleva a su mano derecha a pellizcar su brazo izquierdo. Al hacerlo, usted sentirá un dolor leve. Ese dolor tiene las siguientes características, más o menos evidentes. Sólo existe en cuanto se lo experimenta de manera consciente y, en consecuencia, es en un sentido de la palabra completamente "subjetivo" y no "objetivo". Por otra parte, hay cierta sensación cualitativa. Así, el dolor consciente tiene al menos estas dos características: subjetividad y cualitatividad.

Pretendo que todo esto suene bastante inocente e incluso aburrido. Hasta aquí, el lector ha tenido tres tipos de experiencia consciente: pensar en algo, hacer algo de manera intencional y tener una sensación. ¿Cuál es el problema? Bien, ahora mire los objetos a su alrededor, las sillas y las mesas, las casas y los árboles. Estos objetos no son "subjetivos" en ningún sentido. Existen con completa independencia de que se los experimente o no. Además, sabemos por otro lado que están hechos en su totalidad de las partículas descritas por la física atómica y que la sensación producida por una partícula física o, pongamos por caso, una mesa no tiene carácter cualitativo. Son partes del mundo que existen al margen de las experiencias. Ahora bien, este sencillo contraste entre nuestras experiencias y el mundo existente con independencia de ellas invita a hacer una caracterización; en nuestro vocabulario tradicional, la caracterización más natural es decir que hay una distinción entre lo mental, por un lado, y lo físico o material, por otro. Lo mental como tal no es físico. Y lo físico como tal no es mental. Esta simple imagen conduce a muchos de los problemas, tres de los cuales, quizá los peores, son ilustrados por nuestros tres ejemplos de apariencia inofensiva. ¿Cómo puede una experiencia consciente como el dolor existir en un mundo que está

íntegramente compuesto de partículas físicas, y cómo pueden algunas de estas, cuya presunta localización es nuestro cerebro, causar las experiencias mentales? (Este es el denominado "problema mente-cuerpo"). Pero aun si llegáramos a una solución de este problema, no estaríamos todavía libres de preocupaciones, pues la siguiente pregunta obvia es: ¿cómo pueden los estados mentales de conciencia subjetivos, insustanciales y no físicos causar algo en el mundo físico? ¿Cómo puede nuestra intención, que no forma parte del mundo físico, causar el movimiento de nuestro brazo? (Este es el llamado "problema de la causación mental"). Por último, los pensamientos del lector en torno de cuestiones políticas plantean un tercer problema inabordable. ¿Cómo pueden esos pensamientos, presuntamente situados en la cabeza, referirse o vincularse a objetos y situaciones distantes, sucesos políticos que ocurren, por ejemplo, en Washington, Londres o París? (Este es el llamado "problema de la intencionalidad", donde "intencionalidad" alude a la facultad direccional o referencial [*aboutness*] de la mente)*.

Nuestras inocentes experiencias invitaban a una descripción; y el vocabulario tradicional de lo "mental" y lo "físico" es difícil de resistir. Ese vocabulario supone la exclusión mutua de lo uno y lo otro, y el supuesto genera problemas insolubles que suscitaron la aparición de un millar de libros. Las personas que aceptan la realidad e

* *Aboutness* no tiene en español una traducción que pueda calificarse de canónica. Se han propuesto, entre otros, términos como "referencialidad", "tendencialidad", "acerqueidad" e incluso "intencionalidad", que no corresponde utilizar aquí porque el autor emplea de manera específica la palabra *intentionality*. Sea como fuere, debe entenderse que *aboutness* alude a la cualidad de la mente de "referirse a" algo. (N. del T.)

irreductibilidad de lo mental tienden a verse a sí mismas como dualistas. Pero para otros, el hecho de aceptar un componente mental irreductible en la realidad se asemeja a renunciar a la cosmovisión científica, por lo cual niegan la existencia de esa realidad mental. Creen que esta puede reducirse a lo material o eliminarse por completo. Y suelen autocalificarse de materialistas. Me parece que unos y otros cometen el mismo error.

Voy a tratar de superar ese vocabulario y sus supuestos, y al hacerlo intentaré resolver o disolver los problemas tradicionales. Pero una vez hecho esto, el tema, la filosofía de la mente, no se acabará: será más interesante. Y esa es la segunda razón por la cual quiero escribir este libro. La mayoría de las introducciones generales al tema se refieren sólo a las Grandes Preguntas. Se concentran sobre todo en el problema mente-cuerpo y también dedican cierta atención al problema de la causación mental y un poco menos al de la intencionalidad. A mi juicio, estas no son las únicas cuestiones interesantes de la filosofía de la mente. Hechas a un lado las grandes preguntas, podemos responder un conjunto de cuestiones más interesante e ignorado: ¿cómo trabaja la mente en detalle?

De manera específica, me parece necesario investigar las cuestiones sobre la estructura detallada de la conciencia y la significación de las recientes investigaciones neurobiológicas sobre el tema. Dedico todo un capítulo a estos asuntos. Respondido el enigma filosófico acerca de la posibilidad de la intencionalidad, podremos seguir adelante y examinar la estructura concreta de la intencionalidad humana. Por otra parte, hay una serie de cuestiones absolutamente fundamentales que debemos aclarar antes de comenzar siquiera a suponer que entendemos el funcionamiento de la mente. Esas cuestiones abarcan más de lo que puedo abordar en un solo libro, no obstante lo cual

consagro un capítulo a cada uno de los siguientes temas: el problema de la libertad de la voluntad, el modo real de operar de la causación mental, la naturaleza y el funcionamiento del inconsciente, el análisis de la percepción y el concepto del yo. En un libro introductorio no puedo abundar en detalles, pero sí proporcionar al menos una idea de la riqueza del tópico, una riqueza que se pierde en los tratamientos habituales del tema en los textos introductorios.

Es preciso trazar con claridad dos distinciones desde el inicio, porque son esenciales para el argumento y porque los malentendidos generados al respecto han provocado una masiva confusión filosófica. La primera es la distinción entre los rasgos de un mundo que son independientes del observador y los que dependen de este o son relativos a él. Considérense las cosas que existirían con independencia de lo que los seres humanos pensaran o hicieran. Algunas de esas cosas son la fuerza, la masa, la atracción gravitatoria, el sistema planetario, la fotosíntesis y los átomos de hidrógeno. Todas ellas son independientes del observador en el sentido de que su existencia no depende de actitudes humanas. Pero hay muchas cosas cuya existencia depende de nosotros y de nuestras actitudes. El dinero, las propiedades, el gobierno, los partidos de fútbol y los cocteles son lo que son, en gran parte, porque eso es lo que pensamos que son. Todas ellas son relativas al observador o dependientes de él. En general, las ciencias naturales se ocupan de los fenómenos independientes del observador y las ciencias sociales abordan los que dependen de este. Los hechos dependientes del observador son creados por agentes conscientes, pero los estados mentales de quienes los crean son, en sí mismos, hechos independientes del observador. Así, el pedazo de papel que tengo en las manos sólo es dinero porque otros

y yo lo consideramos como tal. El dinero es dependiente del observador. Pero el hecho mismo de que lo consideremos dinero no depende del observador. Que otros y yo le demos ese carácter es un hecho acerca de nosotros e independiente del observador.

En lo concerniente a la mente también debemos distinguir entre intencionalidad original o intrínseca, por una parte, e intencionalidad derivada, por otra. Por ejemplo, en la cabeza tengo *información* sobre la manera de llegar a San José. Tengo un conjunto de *creencias* verdaderas acerca del camino a esa localidad. Esa información y esas creencias presentes en mí son ejemplos de intencionalidad original o intrínseca. El mapa frente a mí también contiene información sobre el modo de llegar a San José, así como ciertos símbolos y expresiones que se *refieren a*, *versan sobre* o *representan* ciudades, autopistas y cosas por el estilo. Pero si el mapa contiene intencionalidad en forma de información, referencialidad y representaciones, lo hace en un sentido derivado de la intencionalidad original de cartógrafos y usuarios. Intrínsecamente, el mapa es sólo una lámina de fibra de celulosa con manchas de tinta. Cualquiera sea su intencionalidad, le es impuesta por la intencionalidad original de los seres humanos.

Es preciso, entonces, tener presentes dos distinciones: en primer lugar, entre los fenómenos dependientes e independientes del observador, y, en segundo lugar, entre la intencionalidad original y la intencionalidad derivada. Su relación es sistemática: la intencionalidad derivada siempre es dependiente del observador.

CAPÍTULO I

UNA DOCENA DE PROBLEMAS DE LA FILOSOFÍA DE LA MENTE

La meta de este libro es introducir al lector en la filosofía de la mente. Mis objetivos son tres. En primer lugar, el lector debe poder comprender las cuestiones y discusiones contemporáneas más importantes en este campo, y también alcanzar cierta comprensión de sus antecedentes históricos. Segundo, quiero dejar establecido con claridad el camino correcto, a mi juicio, para abordar estos problemas, e incluso espero dar respuestas a muchos de los interrogantes que planteo. En tercer lugar, lo más importante: me gustaría que el lector pudiera pensar por sí mismo estas cuestiones luego de leer el libro. Puedo enunciar de una vez las tres metas si digo que trato de escribir el libro que querría haber leído cuando comencé a reflexionar sobre estos asuntos. Escribo con la convicción de que la filosofía de la mente es el tema más importante de la filosofía contemporánea, y que las visiones corrientes –dualismo, materialismo, conductismo, funcionalismo, computacionalismo, eliminativismo, epifenomenalismo– son falsas.

Una de las cosas agradables de escribir sobre la mente es que no hace falta explicar por qué el tema es importante. Es preciso algún tiempo para comprender la importancia filosófica de los actos ilocutivos y la lógica modal cuantificada, pero todo el mundo advierte de inmediato que la mente tiene un papel central en nuestra vida. Su funcionamiento –consciente e inconsciente; libre y no libre; en la percepción, la acción y el pensamiento; en los sentimientos, la emoción, la reflexión y la memoria y en todos sus otros rasgos– no es tanto un

aspecto de nuestra vida como, en cierto sentido, nuestra vida misma.

Al escribir un libro de este tipo se corren algunos riesgos: una de las peores cosas que podemos hacer es suscitar en los lectores la impresión de que entienden algo que en realidad no entienden, de que algo ha sido explicado cuando no es así y de que se ha resuelto un problema cuando no se ha encontrado su solución. Tengo aguda conciencia de esos riesgos, y en las páginas que siguen haré hincapié tanto en zonas de la ignorancia humana –la mía y la de otros– como en zonas del entendimiento humano. Creo que la filosofía de la mente es tan importante que vale la pena correr esos riesgos. Por una serie de importantes razones históricas, ese campo se ha convertido en el tópico central de la filosofía contemporánea. Durante la mayor parte del siglo XX la filosofía del lenguaje fue la "filosofía primera". Otras ramas de la filosofía se consideraban derivadas de aquella y dependían de los resultados obtenidos por esta para alcanzar su solución. Hoy, el foco de la atención se ha desplazado del lenguaje a la mente. ¿Por qué? Bien, creo en primer término que muchos de quienes trabajamos en la filosofía del lenguaje vemos numerosas cuestiones lingüísticas como casos especiales de cuestiones referidas a la mente. Nuestro uso del lenguaje es una expresión de nuestras capacidades mentales más fundamentales en términos biológicos, y no entenderemos del todo su funcionamiento mientras no comprendamos que se basa en nuestras habilidades mentales. Una segunda razón es que el desarrollo de los conocimientos nos ha permitido dejar atrás la idea de que la teoría del conocimiento, la epistemología, es central en nuestra disciplina, y ahora estamos preparados para hacer una filosofía constructiva más teórica y

sustantiva, en vez de ocuparnos uno por uno de problemas tradicionales específicos. El lugar ideal para comenzar esa filosofía constructiva es el examen de la naturaleza de la mente humana. Una tercera razón de la centralidad de la mente es que para muchos –yo incluido– la cuestión esencial de la filosofía a principios del siglo XXI es cómo hacer una descripción de nosotros mismos como agentes aparentemente conscientes, atentos, libres, racionales, parlantes, sociales y políticos en un mundo consistente en su totalidad, según nos dice la ciencia, de partículas físicas sin sentido ni significado. ¿Quiénes somos, y cómo encajamos en el resto del mundo? ¿Cómo se relaciona la realidad humana con el resto de la realidad? Una forma especial de esta pregunta es: ¿qué significa ser humano? Para responder a estas preguntas es preciso comenzar con una discusión sobre la mente, porque los fenómenos mentales constituyen el puente a través del cual nos conectamos con el resto del mundo. Una cuarta razón de la preponderancia de la filosofía de la mente ha sido la invención de la "ciencia cognitiva", una nueva disciplina que intenta profundizar en la naturaleza mental más de lo que solía hacerse en la psicología empírica tradicional. La ciencia cognitiva necesita fundarse en la filosofía de la mente. Por último, y de manera más polémica, creo que la filosofía del lenguaje ha alcanzado un período de relativo estancamiento debido a ciertos errores comunes que rodean la doctrina del llamado externalismo, la idea de que el significado de las palabras, y por extensión los contenidos de nuestra mente, no están dentro de la cabeza y tienen que ver, en cambio, con relaciones causales entre lo que hay en ella y el mundo externo. No es este el lugar para exponer esas cuestiones en detalle, pero el hecho de no haber podido presentar una

descripción del lenguaje sobre la base de una premisa externalista ha generado un período de inactividad en la filosofía correspondiente, y la filosofía de la mente ocupó el lugar vacío. Diré más sobre el externalismo en el capítulo 6.

La filosofía de la mente tiene una característica especial que la distingue de otras ramas de la filosofía. En la mayoría de los temas filosóficos no hay una marcada división entre las convicciones de los profesionales y las opiniones del público instruido. Pero en las cuestiones discutidas en este libro existe una enorme diferencia entre lo que cree la mayor parte de la gente y lo que afirman los profesionales expertos. Supongo que en el mundo occidental la mayoría acepta en nuestros días alguna forma de dualismo. Creen que tienen una mente, o un alma, y un cuerpo. Incluso he escuchado a algunas personas decirme que tienen tres partes: un cuerpo, una mente y un alma. Pero sin lugar a dudas no es eso lo que opinan los profesionales de la filosofía, la psicología, la ciencia cognitiva, la neurobiología o la inteligencia artificial. Casi sin excepción, los expertos pertenecientes a esos campos aceptan alguna versión del materialismo. En este libro se dedicará un gran esfuerzo a tratar de explicar esas cuestiones y a resolver los problemas concomitantes.

Supongamos entonces que la mente es hoy el tópico central de la filosofía y que otras cuestiones, como la naturaleza del lenguaje y el significado, la naturaleza de la sociedad y la naturaleza del conocimiento, son de una manera u otra casos especiales de las características más generales de la mente humana. ¿Cómo debemos proceder a examinar la mente?

1. Descartes y otros desastres

En la filosofía no hay escape de la historia. A veces creo que lo ideal sería contar a mis alumnos la verdad sobre una cuestión y despacharlos a su casa. Pero ese enfoque totalmente ahistórico tiende a producir superficialidad filosófica. Debemos saber cómo llegamos históricamente a plantearnos las cuestiones que nos ocupan y qué tipo de respuestas les dieron nuestros antecesores. En la era moderna, la filosofía de la mente comienza en efecto con la obra de René Descartes (1596-1650). Descartes no fue el primero en sostener puntos de vista como los que sostuvo, pero su concepción de la mente fue la más influyente entre las propuestas por los llamados filósofos modernos, los filósofos del siglo XVII y siguientes. Muchas de sus ideas se exponen de manera rutinaria, y gente que ni siquiera puede pronunciar el apellido del filósofo las acepta acríticamente. La doctrina más famosa de Descartes es el dualismo, la idea de que el mundo se divide en dos clases diferentes de *sustancias* o entidades de existencia autónoma. Se trata de las sustancias mentales y las sustancias físicas. A veces, el dualismo cartesiano recibe el nombre de "dualismo sustancial"[1].

Descartes creía que una sustancia debía tener una esencia o un rasgo esencial que la hacía ser lo que era (por cierto, toda esta jerga sobre la sustancia y la esencia proviene de Aristóteles). La esencia de la mente es la conciencia o el "pensamiento", como él la denominó; y la esencia del cuerpo es el hecho de extenderse en tres dimensiones del espacio físico: la "extensión"

1 No pretendo sugerir que la mía es la única interpretación razonable de Descartes. Afirmo, antes bien, que la interpretación aquí presentada ha sido la de mayor influencia en la historia del tema.

en el vocabulario cartesiano. Al decir que la esencia de la mente es la conciencia, Descartes afirma que somos la clase de seres que somos por ser conscientes; siempre nos encontramos en algún estado consciente y dejaríamos de existir si no fuera así. Por ejemplo, en este mismo momento mi mente se concentra conscientemente en la escritura del primer capítulo del presente libro, pero, cualesquiera sean los cambios que yo atraviese cuando deje de escribir y, tomemos por caso, empiece a cenar, permaneceré en un estado consciente. Al decir que la esencia del cuerpo es la extensión, Descartes sostiene que los cuerpos tienen dimensiones espaciales: el escritorio frente a mí, el planeta Tierra y el auto en el estacionamiento se extienden o difunden en el espacio. En la terminología latina de Descartes la distinción es entre *res cogitans* y *res extensa*. (El apellido del filósofo, dicho sea de paso, es una contracción de "Des Cartes"; en latín: "Cartesius", que significa "de las cartas", y el adjetivo correspondiente en español es "Cartesiano".)

El dualismo cartesiano fue importante en el siglo XVII por varias razones, sobre todo porque parecía dividir el territorio entre ciencia y religión. Los nuevos descubrimientos científicos hechos durante esa centuria parecían plantear una amenaza a la religión tradicional y había terribles disputas sobre el conflicto aparente entre fe y razón. Aunque no por completo, Descartes desactivó este conflicto al asignar, de hecho, el mundo material a los científicos y el mundo mental a los teólogos. La mente se concebía como un alma inmortal y no era un tópico apropiado para las indagaciones científicas, mientras que los cuerpos podían ser investigados por ciencias como la biología, la física y la astronomía. La filosofía, por cierto, podía a juicio de Descartes estudiar tanto la mente como el cuerpo.

Según nuestro filósofo, cada esencia tiene diferentes modos o modificaciones en los cuales puede manifestarse. Los cuerpos son infinitamente divisibles. Es decir, pueden en principio dividirse de manera indefinida en partes más pequeñas, y en ese sentido es posible destruir un cuerpo, pero no la materia en general. La cantidad de materia existente en el universo es constante. Las mentes, por su parte, son indivisibles, esto es, no se las puede dividir en partes más pequeñas, por lo cual es imposible destruirlas como los cuerpos. Cada mente es un alma inmortal. Los cuerpos, en cuanto entidades físicas, están determinados por las leyes de la física; las mentes, en cambio, tienen libre albedrío. Cada uno de nosotros, en cuanto yo, es idéntico a su mente. En cuanto seres humanos vivos somos entidades compuestas con una mente y un cuerpo, pero para cada uno de nosotros el yo [*self*], el objeto al que hacemos referencia con la palabra "yo" ["*I*"], es una mente atada de algún modo a nuestro cuerpo. Gilbert Ryle, un filósofo de la mente del siglo XX, se mofaba de este aspecto de la concepción de Descartes y lo llamaba la doctrina del "fantasma en la máquina". Cada uno de nosotros es un fantasma (nuestra mente) que habita una máquina (nuestro cuerpo)[2]. Conocemos la existencia y los contenidos de nuestra mente en virtud de una especie de conocimiento inmediato, que Descartes compendia en la sentencia más célebre de su filosofía, "*cogito ergo sum*": pienso, luego existo. La máxima se parece a un argumento formal en el cual "pienso" es la premisa y "existo" la conclusión, pero creo que Descartes pre-

2 G. Ryle, *The Concept of Mind*, Londres, Hutchinson, 1949 [traducción española: *El concepto de lo mental*, Buenos Aires, Paidós, 1967].

tendía que también indicara una suerte de inspección interna de la existencia y los contenidos de la mente. No puedo equivocarme en lo concerniente a la existencia de mi propia conciencia, por ende no puedo equivocarme acerca de mi existencia, porque mi esencia consiste en ser consciente (es decir, pensante), en ser una mente. Tampoco puedo errar con respecto a los contenidos de mi mente. Si me parece, por ejemplo, que tengo un dolor, sin duda lo tengo.

Los cuerpos, por su parte, no pueden conocerse directamente sino de manera indirecta, deduciendo su existencia y características a partir de los contenidos de la mente. No percibo directamente la mesa frente a mí; estrictamente hablando, sólo percibo mi experiencia consciente de la mesa, mi "idea" de ella, e infiero su existencia de la presencia de esta última. Mi idea presente de la mesa no es causada por mí, razón por la cual debo suponer que su causa es la propia mesa.

La descripción cartesiana de la relación entre mente y cuerpo puede resumirse en el siguiente cuadro. Además de tener una esencia, cada sustancia tiene una serie de modificaciones o propiedades, y estas son las formas particulares adoptadas por la esencia.

	Sustancias	
	Mente	Cuerpo
Esencia	Pensamiento (conciencia)	Extensión (posee dimensiones espaciales)
Propiedades	Conocida directamente Libre Indivisible Indestructible	Conocido indirectamente Determinado Infinitamente divisible Destructible

Las concepciones cartesianas han suscitado debates interminables, y es justo decir que Descartes nos

legó más problemas que soluciones. Por breve que sea, la descripción que acabo de presentar, la de la realidad dividida entre lo mental y lo físico, nos plantea un quintal de problemas. Ocho de ellos fueron los que más preocuparon al propio Descartes y a sus sucesores inmediatos, y quiero examinarlos a continuación.

1. El problema mente-cuerpo

¿Cuáles son exactamente las relaciones entre lo mental y lo físico? En particular, ¿cómo puede haber relaciones causales entre ellos? Parece imposible que pueda haberlas entre dos reinos metafísicos completamente diferentes, el reino físico de los objetos materiales extensos y el reino mental o espiritual de las mentes o almas. ¿Cómo es que algo perteneciente al cuerpo puede causar algo en la mente? ¿Cómo algo perteneciente a la mente causa algo en el cuerpo? No obstante, al parecer sabemos que hay relaciones causales. Sé que si alguien me pisa el pie, siento un dolor aun cuando el pisotón sólo sea un suceso físico del mundo físico, mientras que mi sensación de dolor es un suceso mental ocurrido en mi alma. ¿Cómo pueden suceder esas cosas? Peor: al parecer también hay relaciones causales en el otro sentido. Decido levantar el brazo, un hecho que ocurre dentro de mi alma consciente, y hete aquí que el brazo se eleva. ¿Cómo debemos concebir que esas cosas puedan siquiera pasar? ¿Cómo puede una decisión de mi alma causar un movimiento en un objeto físico del mundo como es mi cuerpo? Este es el más famoso problema legado por Descartes, y suele llamárselo "problema mente-cuerpo". ¿Cómo puede haber relaciones causales entre una y otro? Gran parte de la filosofía de la mente posterior a Descartes se ocupa de

este problema, que es, a pesar de todos los progresos realizados a través de los siglos, una de las principales cuestiones de la filosofía contemporánea. A mi entender, tiene una solución filosófica general bastante obvia, que explicaré más adelante; pero debo anticipar que muchos de mis colegas –acaso la mayoría– manifiestan un enérgico desacuerdo con mi afirmación de que podemos dar una rápida solución al problema de Descartes.

En realidad, hay dos conjuntos de problemas. Cómo puede algo físico producir un efecto en mi alma, que no es física, y cómo pueden los sucesos de mi alma afectar el mundo físico. En los últimos ciento cincuenta años la primera de esas preguntas se reformuló de una manera que Descartes no habría aceptado. En su versión moderna, reza así: ¿cómo pueden los procesos cerebrales producir fenómenos mentales? ¿Cómo puede el cerebro ser la causa de la mente? Descartes no creía que eso fuera posible, porque de acuerdo con su descripción las mentes tienen una existencia completamente independiente del cerebro. Para él, el problema no era la cuestión *general* de cómo puede surgir una sustancia de la neurobiología, porque a su juicio eso no podía suceder. Antes bien, se preguntaba cómo, a raíz de una herida en mi cuerpo, pueden surgir contenidos mentales *específicos* como una sensación de dolor. Suponemos que la existencia misma de una mente se explica por las operaciones del cerebro. Descartes no creía que eso fuera posible. La cuestión, a su entender, sólo era cómo pueden pensamientos y sentimientos *específicos*, como una sensación de dolor, ser causados por hechos ocurridos al cuerpo.

Es importante destacar este punto: tendemos a creer, y lo hacen aun quienes son dualistas, que nuestro cuerpo, con su cerebro, es consciente. Descartes no

compartía esa creencia. Consideraba que el cuerpo y el cerebro no podían ser más conscientes que las mesas, las sillas, las casas o un montón cualquiera de chatarra. El alma consciente está separada del cuerpo humano, aunque de alguna manera permanece unida a él. Pero ningún objeto material, vivo o muerto, es consciente.

2. *El problema de la existencia de otras mentes*

Dije que, de acuerdo con Descartes, cada uno de nosotros es una mente y conoce los contenidos de esta en forma directa; ¿cómo sé, empero, que otras personas tiene mente? ¿Qué me hace estar seguro, por ejemplo al encontrarme contigo, de que tienes una mente? A fin de cuentas, todo lo que puedo percibir es tu cuerpo, incluyendo su movimiento físico y los sonidos que salen de su boca, que interpreto como palabras. Pero ¿cómo sé que hay algo detrás de todos esos fenómenos físicos? ¿Cómo sé que tienes una mente, cuando la única de la cual tengo un conocimiento directo es la mía?

Cabría estimar que puedo inferir la existencia de estados mentales en ti por analogía conmigo mismo. Así como en mi caso observo una correlación entre estímulo entrante, estado mental interno, y comportamiento de salida, en el tuyo, al advertir el estímulo entrante y el comportamiento de salida, infiero por analogía que debes tener un estado mental interno correspondiente al mío. Así, si me golpeo el pulgar con un martillo, el estímulo entrante me hace sentir dolor, lo cual me lleva a su vez a gritar. En tu caso –así dice la historia–, observo el estímulo entrante y el grito, y simplemente completo el proceso haciendo una analogía entre tú y yo.

Este es un famoso argumento, llamado "argumento de la analogía". Pero no funciona. En general, una de las exigencias hechas al conocimiento inferencial es que, para ser válido, debe haber en principio alguna manera independiente o no inferencial de verificar la inferencia. Así, si creo que hay alguien en la habitación de al lado porque los sonidos que oigo me hacen inferir su presencia, siempre cabe la posibilidad de ir a verificar la inferencia y constatar que, en efecto, hay en ese cuarto alguien que los causa. Pero si sobre la base de tu estímulo y tu comportamiento hago una inferencia sobre tu estado mental, ¿de qué manera puedo verificarla? ¿Cómo puedo acaso comprobar que infiero correctamente y no que sólo formulo una conjetura infundada? Si supongo que el hecho de que tengas o no estados mentales correspondientes a tus estímulos observables y tus patrones de respuesta –así como yo tengo los míos, correspondientes a mis estímulos y patrones– es una hipótesis científica que debemos verificar mediante métodos científicos, el argumento prueba, al parecer, que soy la única persona en el mundo que tiene algún estado mental. Asi, por ejemplo, si pido a todos los presentes en la habitación que pongan los pulgares sobre un escritorio y voy golpeándolos con un martillo para ver si a alguien le duele, resulta ser que, hasta donde puedo constatar, sólo hay uno que duele: el pulgar que llamo mío, pues cuando golpeo los demás no hay sensación alguna.

La concepción de que soy la única persona que tiene estados mentales se denomina "solipsismo". El solipsismo tiene al menos tres grados. Uno, la forma más extrema: soy la única persona en el mundo que tiene estados mentales; y en algunas versiones, nada existe en el mundo salvo mis estados mentales. Dos, el

solipsismo epistémico: tal vez otras personas tengan estados mentales, pero nunca puedo saberlo con certeza. Es muy posible que los tengan, pero no tengo manera de comprobarlo, porque todo lo que puedo observar es su comportamiento externo. Y tres: otras personas tienen estados mentales, pero jamás puedo estar seguro de que son como los míos. Por lo poco que sé, si tú pudieras tener la misma experiencia que yo llamo "ver rojo", quizá la llamarías "ver verde", y si yo tuviera la experiencia que denominas "ver rojo", la llamaría "ver verde". Ambos pasamos las mismas pruebas de daltonismo porque ambos hacemos las mismas discriminaciones en nuestro comportamiento. Si se nos pide que tomemos el lápiz verde de una caja de lápices rojos, los dos tomamos el mismo. Pero ¿cómo sé que tus experiencias internas, que te permiten discriminar, son similares a las que me permiten hacer otro tanto?

El solipsismo es infrecuente en la historia de la filosofía, por cuanto no hay solipsistas célebres. Prácticamente no hubo posición filosófica imaginable, por loca que fuera, que no haya sido sostenida por uno u otro filósofo célebre, pero, por lo que sé, ningún filósofo de fama histórica ha sido jamás solipsista. Desde luego, si alguien lo fuera, difícilmente perdería tiempo en decirnos que lo es, porque de acuerdo con su propia teoría no existimos[3].

3 Bertrand Russell escribe: "Contra el solipsismo debe decirse, en primer lugar, que es psicológicamente imposible de creer, y de hecho es rechazado aun por quienes pretenden aceptarlo. Una vez recibí una carta de una eminente especialista en lógica, la señora Christine Ladd Franklin, en la que me decía que era solipsista y estaba sorprendida de que no hubiera otros". Véase B. Russell, *Human Knowledge: Its Scope and Limits*, Londres, Allen and Unwin, 1948, p. 180 [traducción es-

Esta doctrina también implica una peculiar asimetría, en cuanto tu solipsismo no es una amenaza para mí, y el mío, si sintiera la tentación de ser solipsista, no podrías refutarlo. Así, por ejemplo, si alguien viene y me dice: "Soy solipsista. Tú no existes", no me sentiré tentado a pensar "¡Dios! Tal vez tenga razón, quizá no existo". Pero, a la inversa, si el solipsismo es mi afición, es inútil que vaya y le diga: "¿Existes? ¿Tienes realmente estados mentales?", porque todo lo que él diga seguirá siendo congruente con la hipótesis solipsista.

3. El problema del escepticismo con respecto al mundo externo, y 4. El análisis de la percepción

El escepticismo acerca de la existencia de otras mentes que se sigue del dualismo cartesiano es sólo un caso especial de una clase de escepticismo mucho más general: el que cuestiona la existencia del mundo externo. Según el punto de vista de Descartes sólo puedo tener un conocimiento cierto de los contenidos de mi mente, mis pensamientos, sentimientos y percepciones concretas, etc. Pero ¿qué pasa con las sillas, las mesas, las montañas, los ríos, los bosques y los árboles que veo a mi alrededor? ¿Tengo un conocimiento cierto de que existen en la realidad y los percibo tal y como efectivamente son? Es importante tener en cuenta que, de acuerdo con la concepción cartesiana, no percibimos directamente los objetos y situaciones del mundo. Lo que percibimos directamente, es decir, sin ningún proceso inferencial, son los contenidos de nuestra mente. Así pues, si sostengo la mano frente a la cara, lo que percibo en forma directa, lo que percibo estricta y literalmente, según Descartes, es cierta experiencia visual. Descartes da el nombre de "ideas" a esas experiencias. No percibo la mano en sí, sino una representación visual determinada de la mano, una especie de imagen mental de ella. Pero entonces surge el siguiente interrogante: ¿cómo sé que frente a mí hay efectivamente una mano que me lleva a tener esa imagen mental? Como no percibo la mano en sí misma sino una representación mental de la mano, debo preguntarme: ¿cómo sé que la representación representa realmente, o con exactitud? El punto de vista de Descartes era corriente en el siglo XVII. Se lo llamaba "teoría representativa de la percepción"; más adelante diré algo más sobre ella, pero en este punto quiero señalar que para Descartes uno de los problemas es: ¿cómo podemos estar realmente seguros, cómo podemos tener un conocimiento cierto y seguro de la existencia de un objeto que genera en mí esa experiencia visual, y de que esta es en todo respecto una representación precisa de las características reales del objeto?

Descartes hace muy poco en términos argumetativos para demostrar que no podemos percibir directamente mesas, sillas, montañas, etc., y que sólo percibimos nuestras ideas de esas cosas. La transición de la percepción de objetos reales a la percepción exclusiva de los contenidos de nuestra mente se da en él de manera muy casual. Aunque no era en modo alguno el primer filósofo en sostener ese punto de vista, el paso de la concepción de que percibimos efectivamente objetos reales a la concepción de que sólo percibimos nuestras ideas de los objetos es un movimiento de decisiva importancia en la historia de la filosofía. En rigor, yo diría que

pañola: *El conocimiento humano: su alcance y sus límites*, Madrid, Taurus, 1977].

es el mayor desastre en la historia de la filosofía a lo largo de los últimos cuatro siglos. En la jerga contemporánea esta perspectiva se expresa así: no percibimos objetos materiales, sólo percibimos "datos de los sentidos" [*sense data*]. Diré mucho más sobre esta cuestión en el capítulo 10.

En realidad, hay dos problemas íntimamente relacionados. El primero es: ¿cómo presentamos un análisis de nuestras interacciones perceptivas con el mundo? ¿Cuál es la relación precisa entre nuestras experiencias perceptivas internas, por un lado, y los objetos materiales y otros rasgos del mundo externo, por otro? El segundo es: ¿cómo podemos estar seguros de que tenemos conocimiento de un mundo externo que está del otro lado de nuestras experiencias perceptivas? Ambos problemas están estrechamente relacionados entre sí porque nos gustaría que nuestro análisis de la percepción del mundo externo nos proporcionara herramientas para refutar el escepticismo acerca de la posibilidad de tener conocimiento de dicho mundo.

5. *El problema del libre albedrío*

He pasado por la experiencia de elegir una entre varias opciones, de decidir entre alternativas genuinas y hacer una cosa cuando podría con toda facilidad haber hecho otra. Se trata de manifestaciones de lo que considero la libertad de mi voluntad. Pero es natural que surja entonces una pregunta: ¿tengo un auténtico libre albedrío o este sólo es una ilusión? El interrogante se plantea de forma especialmente apremiante para Descartes, porque si el libre albedrío es un rasgo de mi mente, ¿cómo puede tener algún efecto sobre el mundo físico, que está íntegramente determinado? Aunque se

trata de una extensión del problema mente-cuerpo, no es una cuestión igual a este. Aun cuando tuviéramos una solución para dicho problema, aun cuando yo pudiese mostrar que mis pensamientos y sentimientos son capaces de mover mi cuerpo, persistiría la cuestión: ¿cuál es la congruencia de esto con la concepción de la física en la época de Descartes, la visión del mundo físico como un sistema completamente cerrado y determinista en términos causales? Cualquier suceso del mundo físico está determinado por sucesos físicos anteriores. Entonces, aunque pudiéramos demostrar de alguna manera que tenemos libre albedrío mental, esto sería indiferente para el comportamiento de mi cuerpo, porque ese comportamiento es causado por los estados previos de mi cuerpo y del resto del universo físico. El problema del libre albedrío parece arduo para cualquiera, pero plantea dificultades excepcionales para quien acepta el dualismo.

Este problema todavía nos persigue en una forma tan apremiante como en los tiempos de Descartes. Hoy creemos que la física cuántica ha mostrado una indeterminación en el comportamiento de las partículas en el nivel subatómico. No todo está, pues, determinado de la manera supuesta por la física clásica. Pero eso no parece servir de ayuda con el problema del libre albedrío, porque la forma de la indeterminación cuántica es el azar, y azar no es lo mismo que libertad. El hecho de que las partículas del micronivel no estén totalmente determinadas y, por lo tanto, sólo se pueda predecir su comportamiento de manera estadística y no con completa certeza, parece no dar respaldo alguno a la idea de que nuestros actos en apariencia libres lo son en efecto. Aun cuando nuestro proceso de toma de decisiones heredara de algún modo la indeterminación de

los sucesos de nivel cuántico en el cerebro, no tendríamos con ello libre albedrío, sino un elemento aleatorio impredecible en nuestras decisiones y conductas. Diré más al respecto en el capítulo 8.

6. *El yo y la identidad personal*

Hay otro problema para el cual los seguidores de Descartes creyeron que su descripción brindaba una respuesta concluyente, aun cuando él mismo no lo abordó en forma directa: el problema de la existencia del yo y su identidad a través del tiempo y el cambio. Para ver en qué consiste, consideremos el siguiente ejemplo: en este mismo momento me ocupo de una serie de cuestiones mientras contemplo un lago en Suecia. Un mes atrás trabajaba en problemas relacionados mientras miraba el océano en las costas de California. Las experiencias son muy diferentes, pero creo que ambas me pertenecen. ¿Por qué? ¿Con qué justificación? En verdad hay aquí toda una serie de cuestiones, una maraña de filosofía. ¿Qué cosa en esas experiencias las hace experiencias de la misma persona, y qué cosa en mí hace que yo sea la misma persona que era en California? Es tentador decir que una y otra personas son la misma porque ambas tienen el mismo cuerpo. Pero ¿es ese cuerpo realmente esencial para mi identidad? Parece al menos posible imaginar que, como Gregor Samsa en el relato de Kafka, yo podría despertarme en un cuerpo absolutamente distinto. Pero si lo que me hace ser yo no es el mismo cuerpo, ¿qué es? ¿Cuál es la relación entre mi identidad personal y mi identidad corporal? Además de esta o aquella experiencia en particular, ¿tengo también la experiencia de mí mismo como un yo?

La respuesta de los dualistas a estos interrogantes es rápida. Mi cuerpo no tiene nada que ver con mi identidad. Esta consiste por entero en la continuación de la misma sustancia mental, la misma alma o *res cogitans*. Los objetos materiales van y vienen y lo mismo puede pasar con las experiencias, pero mi identidad está garantizada por la mismidad de mi sustancia mental, pues yo soy idéntico a esta.

Para Descartes hay otros dos problemas cuya naturaleza se asemeja más a la de un enigma que debe resolver, pero sus soluciones son muy interesantes. Hablo del problema de los animales no humanos y del problema del sueño.

7. *¿Tienen mente los animales?*

Si toda mente es una sustancia espiritual o mental y las mentes son indestructibles, debe deducirse que si los animales la tienen, todo animal posee un alma inmortal. Pero si cualquier perro, gato, ratón, pulga o saltamontes tiene un alma inmortal, el paraíso, por no decir algo peor, va a estar superpoblado. La solución de Descartes para el problema de la mente animal fue rápida y brutal. Dijo que los animales no tienen mente. Con todo, no se mostraba dogmático al respecto; acaso la tuvieran, pero le parecía científicamente improbable que así fuera. A su juicio, la distinción crucial entre nosotros y ellos, la distinción que nos permite decir con toda seguridad que los seres humanos tienen mente y los animales no, es que los primeros poseen un lenguaje en el cual expresan sus pensamientos y sentimientos, mientras que los segundos no lo tienen. Esa falta de lenguaje era para él una prueba abrumadora de que carecen de pensamientos y sentimientos. Descar-

tes concedía que este es en cierto modo un resultado contrario a la intuición. Si vemos a un perro golpeado por un carro y oímos sus aullidos de aparente dolor, debemos suponer que tiene sentimientos como nosotros. Pero Descartes dice que todo esto es una ilusión. No deberíamos sentir más lástima por el perro de la que sentimos por el carro cuando interviene en el choque. El ruido tal vez nos haga creer que el carro sufre dolor, pero no es así; lo mismo sucede con los perros y los restantes animales. Parece una locura negar que los perros y otros animales sean conscientes, pero esta es la idea que, a mi entender, Descartes tenía de la materia. En el caso humano, el cuerpo no es consciente. Sólo lo es el alma inmortal, que está unida al cuerpo. En cuanto al perro, empero, parece muy improbable que haya un alma inmortal; sólo hay un cuerpo, y los cuerpos no pueden ser conscientes. Por lo tanto, el perro no es consciente. Y lo mismo vale para todos los demás animales.

8. El problema del sueño

El octavo problema de Descartes es el sueño. Si toda mente es esencialmente consciente, esto es, si la conciencia es su esencia, de modo tal que no podríamos tener mente sin ser conscientes, la inconsciencia implicará, al parecer, la no existencia. Y en verdad, esto es lo que da a entender la teoría de Descartes: si dejo de ser consciente, dejo de existir. Pero ¿cómo explicamos entonces el hecho de que la gente, a pesar de estar viva, a menudo esté inconsciente, por ejemplo al dormir? La respuesta de Descartes sería que nunca estamos un ciento por ciento inconscientes. Siempre hay un nivel mínimo de actividad onírica aun en el sueño

más profundo. Mientras sigamos existiendo seguiremos siendo necesariamente conscientes.

II. Cuatro problemas más

De los problemas del ajuste entre la mente y el resto del universo se desprenden otros cuatro problemas que, sin embargo, no fueron abordados por el propio Descartes, o que en la época contemporánea se transformaron de tal manera que presentan una forma muy diferente de la que tenían cuando él y sus sucesores inmediatos se ocuparon de ellos.

9. El problema de la intencionalidad

La intencionalidad es un problema planteado no sólo al dualismo sino a la filosofía de la mente en general. Descartes nunca lo enfrentó de manera explícita, pero en los filósofos ulteriores llegó a ocupar el primer plano y en los últimos cien años se convirtió, a decir verdad, en uno de los problemas centrales de la filosofía de la mente.

"Intencionalidad" es un término técnico utilizado por los filósofos para referirse a la capacidad de la mente en virtud de la cual los estados mentales se refieren a, versan sobre o corresponden a objetos y situaciones del mundo al margen de sí mismos. Así, por ejemplo, si tengo una creencia, debe ser la creencia de que algo es el caso. Si tengo un deseo, debe ser un deseo de hacer algo o de que algo pase. Si tengo una percepción, debo suponer al menos que percibo algo, un objeto o un estado de cosas en el mundo. De todo ello se dice que es intencional, por cuanto en cada caso el estado hace una referencia más allá de sí mismo. La intención

[*intending*], tal cual es de uso corriente cuando digo que tengo la intención de ir al cine esta noche, es sólo un tipo de intencionalidad entre otros, junto con la creencia, la esperanza, el temor, el deseo y la percepción. (El término técnico inglés no procede del *intention* de esa misma lengua, sino del alemán *Intentionalität*, derivado a su vez del latín). Se trata de un término técnico especial que no debe confundirse con la intención en sentido corriente.

El problema filosófico especial de la intencionalidad es el siguiente: supongamos que ahora creo que George W. Bush está en Washington. Surge entonces la pregunta: ¿cómo pueden mis pensamientos, íntegramente localizados en mi mente, llegar hasta Washington D.C.? Si estimo que el Sol está a ciento cincuenta millones de kilómetros de la Tierra, ¿cómo puede ser, otra vez, que mis pensamientos se extiendan hasta él y remitan a una cosa fuera de sí mismos? El interrogante sobre cómo puede un estado mental referirse a o versar sobre algo más allá de sí mismo es el problema de la intencionalidad.

Es absolutamente esencial aclarar la distinción entre la intencionalidad intrínseca u original que tengo en la cabeza cuando pienso en algo y la intencionalidad derivada que tienen las marcas en el papel cuando pongo mis pensamientos por escrito. Las palabras en el papel realmente significan y refieren, por lo cual tienen intencionalidad, pero esta deriva de la mía al escribirlas de manera intencional. También es preciso distinguir estas dos intencionalidades, la original y la derivada, de las atribuciones metafóricas o los casos de "como si" de la intencionalidad. Si tengo sed, estamos ante un caso de intencionalidad intrínseca u original. Si escribo "tengo sed", la frase tiene una intencionalidad derivada. Si

digo "mi auto está sediento de gasolina", la oración hace una atribución metafórica o de "como si" de la sed al automóvil. Pero en un plano literal el auto no tiene ninguna intencionalidad, ni original ni derivada. Es casi imposible decir cuánta confusión ha generado la inadvertencia de estas distinciones elementales.

En su forma moderna la intencionalidad plantea, en realidad, dos problemas. El primero: ¿cómo es posible que los hechos ocurridos en nuestro cerebro remitan más allá de sí mismos? ¿Cómo es siquiera posible la referencialidad o direccionalidad? El segundo está relacionado con el primero: ¿cómo es que nuestro cerebro o nuestra mente tienen los contenidos intencionales específicos que tienen? Así, por ejemplo, si ahora pienso en George W. Bush, ¿qué hecho de mí mismo hace que el contenido de mi creencia se refiera a George W. Bush y no, digamos, a su hermano Jeb o a su padre George Bush, o a otra persona llamada George W. Bush o a mi perro Gilbert? En resumen, los dos problemas se pueden plantear así: ¿cómo es posible la intencionalidad? Y dado que es posible, ¿por qué los estados intencionales tienen los contenidos específicos que tienen? Dedico el capítulo 6 a los problemas de la intencionalidad.

10. Causalidad mental y epifenomenalismo

Dije que el problema mente-cuerpo tenía dos partes, una de entrada y otra de salida. ¿Cómo causan los estímulos entrantes los fenómenos mentales, y cómo causan los fenómenos mentales el comportamiento de salida? Cada uno de estos aspectos merece un examen por separado, de modo que voy a transformar la cuestión del funcionamiento causal de los estados mentales en un tópico independiente.

Algunos filósofos creen posible explicar cómo la conciencia es una causa de los procesos cerebrales, pero no están dispuestos a admitir que tenga poderes causales propios. Se reconoce que, de un modo u otro, la conciencia y los fenómenos mentales en general dependen de procesos cerebrales, pero es difícil comprender cómo podrían causar movimientos corporales o cualquier otra cosa en el mundo físico. La concepción de que los estados mentales existen pero son causalmente inertes se denomina "epifenomenalismo". Según esta perspectiva la conciencia existe, admitido, pero es como la espuma de la ola o el resplandor de la luz solar reflejada en la superficie del agua. Está allí pero realmente no cuenta. Es un epifenómeno. Esto, sin embargo, también parece contrario a la intuición. Cada vez que decido alzar el brazo, este se levanta. Y no se trata de un fenómeno aleatorio o estadístico. No digo: "Bueno, así son las cosas con mi viejo brazo. Algunos días se levanta y otros no". El problema consiste en mostrar que algo que no forma parte del mundo físico puede tener tales efectos sobre este, y en la jerga contemporánea se plantea de la siguiente forma. Con frecuencia se dice: "El mundo físico es causalmente cerrado". Lo cual significa que nada exterior al mundo físico puede entrar a él y actuar de manera causal. ¿Cómo podrían entonces los estados mentales, que no son físicos y por lo tanto no forman parte del mundo físico, actuar causalmente en este?

11. El inconsciente

Para Descartes, toda actividad mental es consciente por definición. La idea de un estado mental inconsciente le parece una contradicción en los términos, una

conciencia inconsciente. Sin embargo, desde hace más o menos un siglo hemos llegado a acomodarnos bastante bien a la idea de que muchos de nuestros estados mentales son inconscientes. ¿Qué puede significar esto? ¿Qué es un estado mental inconsciente? ¿Cómo se ajusta al resto de nuestra vida mental y al mundo en general?

El problema del inconsciente no lo es sólo para la psicopatología. Decimos, en efecto, que la gente actúa por motivos de los cuales son inconscientes y cuya presencia negarían con toda sinceridad. Decimos que Sam insultaba a su hermano Bob porque tenía una hostilidad inconsciente contra él. Este es el tipo de cosas de las que intenta ocuparse la psicología freudiana. Pero hay otro uso más difundido de la noción de inconsciente, según el cual existen numerosos procesos mentales que se desarrollan en el cerebro pero carecen de manifestaciones conscientes. De acuerdo con las teorías convencionales de la percepción, suponemos que los individuos perciben las formas de los objetos infiriendo de manera inconsciente las características reales de estos a partir de los rasgos limitados del estímulo físico que se les presenta. El problema para estas dos concepciones de lo inconsciente es el siguiente: ¿qué significa exactamente este en términos reales? ¿De qué modo podrían los sucesos cerebrales ser a la vez *mentales* e *inconscientes*?

12. Explicación psicológica y social

Las explicaciones de los fenómenos psicológicos y sociales humanos parecen tener una estructura diferente de las explicaciones en la química y la física. Cuando explicamos por qué votamos de tal o cual

modo en las últimas elecciones o por qué estalló la Primera Guerra Mundial, utilizamos al parecer un tipo de explicación diferente del empleado para explicar por qué crecen las plantas. ¿Cuáles son las formas apropiadas de explicación para los fenómenos psicológicos y sociales humanos y qué implicaciones tiene ello para las perspectivas de las ciencias sociales?

Una de las facetas más decepcionantes de la historia intelectual de los últimos cien años fue la imposibilidad de las ciencias sociales de alcanzar el rico poder explicativo característico de las ciencias físicas y biológicas. En sociología, y hasta en economía, carecemos del tipo de estructuras de conocimiento establecidas con que contamos en física y química. ¿Por qué? ¿Por qué los métodos de las ciencias naturales no tuvieron en el estudio del comportamiento y las relaciones sociales humanas la clase de rédito que han tenido en las ciencias físicas?

III. Las soluciones de Descartes a los problemas

Una gran parte de este libro se dedicará a los doce problemas que acabo de esbozar. Si el lector considera interesantes los problemas, es probable que también juzgue interesante el libro. Si no puede imaginarse aunque lo maten por qué alguien habría de interesarse en ellos, con seguridad ha elegido el libro equivocado. Este no es un libro histórico, y no diré demasiado acerca del desarrollo de estos problemas desde el punto de vista de su historia. Sin embargo, puesto que he recurrido a Descartes como fuente para introducir ocho de ellos, quiero contarles, aunque sea brevemente, cuáles fueron

las respuestas del filósofo a estas ocho cuestiones. Creo que esas respuestas fueron inadecuadas sin excepción; para ser justos con él, hay que decir que a menudo fue muy consciente de que lo eran. En mi opinión, el lector entenderá mejor la filosofía contemporánea si ve, al menos en un bosquejo, cómo abordó Descartes estos problemas.

1. El problema mente-cuerpo

En esta cuestión, Descartes nunca alcanzó una respuesta que lo dejara satisfecho. Reconocía que la mente causaba sucesos en el cuerpo y que sucesos de este causaban sucesos en el terreno mental. Pero ¿cuál era exactamente su funcionamiento? Jamás creyó haber resuelto este interrogante. Estudió anatomía y por lo menos una vez observó la disección de un cadáver para tratar de averiguar dónde estaba el punto de conexión entre la mente y el cuerpo. Al final dio con la hipótesis de que debía encontrarse en la glándula pineal, un pequeño órgano en forma de pera situado en la base del cráneo. Descartes suponía que esa glándula era el lugar donde las fuerzas mentales y las fuerzas físicas se ponían en contacto. La idea no es tan alocada como parece: el argumento cartesiano para justificarla era razonable. El filósofo advirtió que todos los elementos cerebrales situados en un lado tenían su réplica en el otro. Debido a la existencia de los dos hemisferios, la anatomía parece mostrarse en duplicado. Pero como todos nuestros sucesos mentales ocurren en forma unitaria, debe haber en el cerebro algún punto unificado donde confluyen las dos corrientes. El único órgano no duplicado que Descartes pudo encontrar dentro del

cerebro fue la glándula pineal, por lo cual supuso que esta debía ser el punto de contacto de lo mental y lo físico.

(El impulso de encontrar el punto de contacto entre el alma y el cuerpo aún persiste. Una vez debatí con un neurobiólogo laureado con el Premio Nobel, sir John Eccles, en la televisión británica. Eccles sostenía que el alma se une al cerebro en el área motriz suplementaria. Este es su argumento: si se pide a un sujeto que lleve a cabo una tarea motriz simple como tocarse cada uno de los dedos de la mano derecha con el pulgar del mismo lado, la corteza motriz muestra un elevado nivel de actividad. Si ahora se le pide que piense la tarea pero no la ejecute, esa misma corteza deja de funcionar, pero el área motriz suplementaria permanece activa. Eccles era de la idea de que cuando sólo el alma está activa, estimula el área motriz suplementaria.)

En un famoso pasaje Descartes dijo que no debíamos imaginar que la mente está alojada en el cuerpo como un piloto en su nave; era preciso pensar que, de algún modo, impregnaba todo el cuerpo. Si tropiezo de frente con algo no observo el choque de mi cuerpo contra otro objeto a la manera como el piloto de un barco podría observar el choque de este contra el muelle; siento, antes bien, dolor en la parte del cuerpo que ha entrado en contacto con el objeto. A juicio de Descartes, debíamos pensar que nuestra mente está en cierta forma difundida a través de todo el cuerpo, pero de acuerdo con su propia doctrina esta afirmación es incorrecta, porque la sustancia mental no puede tener extensión espacial. No puede difundirse a lo largo del cuerpo porque no puede difundirse en absoluto.

2. El problema de la existencia de otras mentes

Con frecuencia se atribuye a Descartes alguna versión del argumento de la analogía, pero yo nunca pude encontrarla explícitamente enunciada en sus escritos. Según dicho argumento, infiero la existencia de estados mentales en otras personas por analogía conmigo mismo. Así como observo una correlación de mi comportamiento con mis estados mentales, puedo inferir la presencia de estados mentales correspondientes en otros al observar su comportamiento. Ya he indicado las limitaciones de esta forma de argumento. El inconveniente del conocimiento inferencial es, en general, que debe haber alguna verificación independiente de la inferencia para que se la considere válida. Así, por ejemplo, yo podría inferir que un recipiente esta vacío si lo golpeara y dedujera por el sonido a hueco que no hay nada en él, pero esta forma inferencial de conocimiento sólo tiene sentido si se supone que puedo abrir el recipiente y mirar en su interior, para constatar de manera no inferencial que, en efecto, está vacío. Sin embargo, en el caso de otras mentes no hay verificación no inferencial de mi inferencia de los estados mentales a partir del comportamiento observado; no hay manera de mirar dentro del recipiente para ver si contiene algo.

3. El escepticismo sobre el mundo externo, y
4. El análisis correcto de la percepción

Por medio de un elaborado argumento, Descartes sostiene que podemos tener un conocimiento cierto de los objetos y situaciones del mundo externo, aun cuando sólo percibimos directamente los contenidos de

nuestra mente. El primer paso de ese argumento exige probar la existencia de Dios, lo cual dista de ser pan comido. No obstante, suponiendo que Dios exista, Descartes aduce que no puede ser engañador. Debido a su perfección, sería incongruente suponer que pudiera serlo, pues el engaño es una imperfección. Pero si Dios no es engañador, debe existir un mundo externo y yo debo tener algún tipo de conocimiento correcto cuando lo observo. ¿Por qué? Porque Dios me da todas las razones para creer, por ejemplo, que hay un escritorio frente a mí y una silla en la cual estoy sentado, y ninguna razón para suponer lo contrario. Por lo tanto, si yo estoy equivocado, Dios me está engañando, lo cual es imposible.

Surge entonces un problema para Descartes: ¿cómo es posible el error? Y su respuesta es que lo es porque mi voluntad excede mi entendimiento. Potencialmente, mi voluntad es infinita; mi entendimiento es finito. Y a menudo quiero creer cosas cuya verdad no percibo de manera clara y distinta; por consiguiente, puedo estar equivocado.

Es importante subrayar que Descartes no creía que nuestras percepciones fueran en general representaciones exactas del mundo. En realidad, los objetos no tienen colores, sabores u olores y tampoco emiten sonidos, pese a que colores, sabores, olores y sonidos nos parecen desde un punto de vista perceptivo partes del mundo. El quid es que podemos estar seguros de que hay un mundo externo causante de nuestras percepciones y gracias a estas podemos obtener algún tipo de información precisa sobre él, aun cuando gran parte de nuestra experiencia perceptiva es ilusoria.

5. *El problema del libre albedrío*

Me parece que, más allá de una mera aserción, Descartes no tiene respuestas para esta cuestión. Dice que soy libre mientras siento que lo soy. Pero el problema, como veremos más adelante, es que no resulta evidente en absoluto que el hecho de percibirme libre indique que lo soy realmente.

6. *El yo y la identidad personal*

Descartes nunca abordó de manera explícita esta cuestión, pero los cartesianos estimaron en general que su dualismo nos da una solución automática al problema. El yo es simplemente idéntico a una sustancia mental y la identidad de esta está garantizada por el mero hecho de ser la misma sustancia mental. Cuesta entender, empero, que la solución propuesta sea otra cosa que una solución por decreto. ¿Cómo llega a adquirir la sustancia mental todos esos misteriosos poderes y propiedades? ¿Y qué razón tenemos para suponer que existe esa sustancia mental además de nuestro cuerpo físico y nuestras experiencias conscientes? Como veremos, David Hume hizo críticas devastadoras de la postura cartesiana sobre el yo y la identidad personal. Según Hume no hay experiencia del yo, y la identidad que nos atribuimos a través de los cambios en nuestra vida es una identidad completamente ficticia: una suerte de ilusión sistemática. Muchos otros filósofos lo siguen en la idea de que no hay nada semejante a un yo por añadidura a la secuencia de nuestras experiencias específicas. Lichtenberg creía que el "yo" ["*I*"] de oraciones como "yo pienso" nos da la ilusión de que hay un "yo" ["*I*"] encargado de pensar; y sostenía, en cam-

bio, que deberíamos decir "eso piensa" ["*it thinks*"], donde el "eso" es tan impersonal como en la frase "llueve" ["*it's raining*"]: en ninguno de estos casos nos referimos concretamente a una entidad.

No hay un solo problema del yo sino varios. No creo que la descripción cartesiana de la *res cogitans* sea en modo alguno una solución a estos problemas, que abordaré en su totalidad en el capítulo 11.

7. Los animales, y 8. El sueño

Ya he criticado las soluciones de Descartes a estos problemas, de modo que seré muy breve: me parece sencillamente descabellado afirmar que los animales no tienen ningún estado consciente. Cuando vuelvo a casa después de trabajar y mi perro viene corriendo a saludarme, moviendo la cola y saltando de un lado a otro, ¿por qué motivo preciso estoy tan seguro de que es consciente y, a decir verdad, que su conciencia tiene un contenido específico, a saber, la alegría de verme? La respuesta habitual a esta pregunta es que, como su comportamiento se asemeja tanto al de una persona feliz, puedo inferir que se trata de un perro contento. Me parece, sin embargo, que este es un argumento erróneo. Para empezar, las personas felices no suelen mover la cola ni tratan de lamerme la mano. Por otra parte –y esto es más importante–, alguien podría construir con toda facilidad un perro robot que moviera la cola y brincara de un lado a otro sin tener absolutamente ningún sentimiento interno. ¿Qué hay de especial en el perro de verdad? A mi entender, la respuesta estriba en que la certeza de que mi perro es consciente y tiene un contenido específico en la conciencia no se funda simplemente en lo apropiado de su comportamiento,

sino en la comprobación de que los fundamentos causales de este son relativamente similares a los míos. Mi perro tiene un cerebro, un aparato perceptivo y una estructura corporal que son notablemente parecidos a los míos: aquí están los ojos, aquí las orejas, aquí la piel, aquí la boca. Si concluyo que es consciente, no lo hago sólo sobre la base de su conducta, sino más bien de la estructura causal que media la relación entre el estímulo entrante y el comportamiento de salida. En el caso de los seres humanos, el estímulo entrante genera experiencias, que a su vez causan un comportamiento de salida. La estructura física subyacente que permite al estímulo entrante causar experiencias es notablemente similar en los humanos y los animales superiores. Por esa razón tenemos completa seguridad de que los perros y los chimpancés tienen estados conscientes, semejantes en muchos aspectos a los nuestros. Cuando se trata de caracoles y termitas, debemos dejar en manos de los expertos la tarea de decirnos si tienen una capacidad neurobiológica lo bastante rica para disfrutar de vida consciente.

Además, así como me parece descabellado suponer que los animales no son conscientes, me parece absurdo suponer que dejamos de existir si tenemos completa inconsciencia durante el sueño o bajo el efecto de la anestesia. Sin embargo, si bien Descartes se equivoca al suponer que la continuación de la conciencia es esencial para la continuidad de nuestra existencia misma, debemos plantear este interrogante: ¿cuáles son exactamente los criterios para determinar que esa existencia tiene continuidad? Nos topamos aquí con el famoso problema de la identidad personal, que analizaré con más detenimiento en el capítulo 11.

Los doce problemas que he esbozado constituyen el marco para mis discusiones sobre la filosofía de la mente. Pero no quiero dar a entender que el tema se limita a ellos. Estos problemas se ramifican en muchos otros que deberemos examinar. Una de las cosas que descubriremos es que con frecuencia hay dos conjuntos de problemas relacionados con cada una de esas cuestiones. Está el problema filosófico inexorable, el "gran problema", por así decirlo, y luego tenemos un problema o grupo de problemas de detalle sobre el funcionamiento de los fenómenos en la vida real. De tal modo, en el caso de la conciencia, por ejemplo, está el gran problema: ¿cómo es siquiera posible una cosa semejante? ¿Cómo *podría* el cerebro causar la conciencia? En los debates actuales se lo suele llamar el "problema duro", y la falta de explicación del papel causal del cerebro recibe el nombre de "laguna explicativa". Pero también hay, me parece, un problema igualmente interesante: ¿cómo funciona la conciencia en organismos concretos como nosotros mismos? Otro tanto puede plantearse con respecto a la intencionalidad. Hay un problema enorme: ¿cómo es posible que la intencionalidad exista? Pero a mi juicio, al menos, la pregunta más interesante es: ¿cómo actúa en detalle?

Lo que he intentado hacer en este capítulo es presentar el marco para los análisis ulteriores. Los problemas no se abordarán como si fueran de igual trascendencia. Ni pensarlo. Los próximos tres capítulos se dedicarán sobre todo al problema mente-cuerpo. Ya he dicho lo que tenía que decir sobre los animales y el sueño. Varios problemas tienen un capítulo propio: la intencionalidad, la causación mental, el libre albedrío, el inconsciente, la percepción y el yo. Algunos de los restantes, aunque son de gran importancia, sólo se harán acreedores a un breve examen en el libro, porque van mucho más allá de la filosofía de la mente; me refiero en especial al escepticismo y la explicación en las ciencias sociales. Se trata de dos grandes cuestiones y sólo las presentaré en un bosquejo, porque para proponer un análisis adecuado haría falta otro libro.

CAPÍTULO 2

EL GIRO HACIA EL MATERIALISMO

1. Dificultades con el dualismo

Damos ahora un salto en el tiempo para trasladarnos a los siglos XX y XXI. Debido a los fracasos del dualismo de estilo cartesiano, en especial su imposibilidad de presentar una descripción adecuada o al menos coherente de la relación entre la mente y el cuerpo, se estima de manera generalizada que el dualismo sustancial ha quedado descartado en cualquiera de sus formas. Esto no significa decir que ningún profesional serio sea partidario de esta doctrina. Según mi experiencia, sin embargo, la mayoría de los dualistas sustanciales que conozco son personas que sostienen esa concepción por motivos religiosos o como parte de una fe religiosa. Una de las consecuencias del dualismo sustancial es que la destrucción de nuestro cuerpo no impide la supervivencia de nuestra alma, por lo cual esta visión es atractiva para los fieles de las religiones que creen en la vida después de la muerte. Pero la mayor parte de los profesionales pertenecientes a este campo no consideran esa doctrina como una posibilidad seria. Una destacada excepción es la defensa del dualismo propuesta por Karl Popper y John C. Eccles[1]. Estos autores afirman la existencia de dos mundos muy distintos, el mundo 1, de los objetos y estados físicos, y el mundo 2, de los estados de conciencia. Se trata de dos mundos separados y distintos que interactúan. En realidad,

1 K. Popper y J. C. Eccles, *The Self and Its Brain*, Berlín, Springer-Verlag, 1977 [traducción española: *El yo y su cerebro*, Barcelona, Labor, 1993].

Popper y Eccles mejoran a Descartes y además postulan un mundo 3, el de "la cultura en todas sus manifestaciones"[2].

Todas las formas del dualismo sustancial heredan el problema cartesiano de cómo hacer una exposición coherente de las relaciones causales entre el alma y el cuerpo, pero versiones recientes presentan un problema adicional. Parece imposible mostrar congruencia alguna entre este dualismo y la física moderna. La física dice que la cantidad de materia/energía en el universo es constante; aquella doctrina, por su parte, parece dar a entender que hay otro tipo de energía, una energía mental o espiritual, no determinada por la física. Así, si el dualismo sustancial es verdadero, debe deducirse, al parecer, que una de las leyes más fundamentales de la física, la ley de la conservación, es falsa. Algunos adeptos de esta corriente han intentado enfrentar el problema afirmando que por cada infusión de energía espiritual hay una disminución de energía física, y de ese modo se preserva una cantidad constante de energía en el universo. Otros han señalado que la mente reordena la distribución de energía universal sin adiciones ni sustracciones. Eccles dice que la mente puede afectar el cuerpo al modificar la probabilidad de sucesos neuronales sin ningún aporte de energía, y agrega que la física cuántica nos permite ver cómo es posible: "La hipótesis de la interacción mente-cerebro es que los sucesos mentales actúan a través de un campo de probabilidad cuántica para alterar la probabilidad de emisión de vesículas de las rejillas vesiculares

2 J. C. Eccles, *How the Self Controls Its Brain*, Berlín, Springer-Verlag, 1994, p. 5.

presinápticas"[3]. En estas maniobras hay algo *ad hoc*, en el sentido de que los autores están convencidos de antemano de la verdad del dualismo y tratan de encontrar alguna manera, cualquiera, de hacerlo compatible con la física.

Es importante entender cuán extrema es la doctrina del dualismo sustancial. Según su perspectiva, nuestro cerebro y nuestro cuerpo no son realmente conscientes. El cuerpo es una mera máquina inconsciente, como un automóvil o un televisor. Está vivo como lo están las plantas, pero en él no hay conciencia. Antes bien, nuestra alma consciente está en cierto modo atada a nuestro cuerpo y seguirá así hasta la muerte de este, en cuyo momento se desprenderá de él. Soy idéntico a mi alma y sólo de manera incidental y temporaria habito este cuerpo.

El inconveniente de esta concepción es que, visto lo que sabemos sobre el funcionamiento del mundo, cuesta tomarla en serio como hipótesis científica. Sabemos que en los seres humanos la conciencia no puede existir en manera alguna sin ciertos procesos físicos que se desenvuelven en el cerebro. Podríamos, en principio, producir conciencia en alguna otra sustancia física, pero por ahora no sabemos cómo hacerlo. Y la idea de que pueda producirse al margen de todo sustrato físico, aunque concebible, parece completamente descartable como hipótesis científica.

No es fácil hacer que la idea de la mente como una sustancia separada sea congruente con el resto de nuestros conocimientos sobre el mundo. A continuación presento tres intentos de hacerlo, cada uno de ellos acorde con una concepción diferente de la mente.

3 Ibíd., p. 69.

Primero, la intervención divina. La ciencia física es incompleta. Nuestra alma es algo que se suma al resto del mundo. Es creada por intervención divina y no forma parte del mundo físico tal como la ciencia lo describe.

Segundo, la mecánica cuántica. El problema mente-cuerpo tradicional sólo se plantea debido a una concepción newtoniana obsoleta de lo físico. Según una interpretación de la medición cuántica, la conciencia es necesaria para completar el colapso de la función de onda y crear así partículas y sucesos cuánticos. De tal modo, hay cierta forma de conciencia que no es creada por el resto de la naturaleza y resulta, en cambio, esencial para la creación de esta última. Es una parte primitiva de la naturaleza requerida para explicar los procesos cerebrales y todo lo demás[4].

Tercero, el idealismo. El universo es enteramente mental. Lo que concebimos como mundo físico es sólo una de las formas adoptadas por la realidad mental subyacente[5].

Menciono estas tres perspectivas para completar el panorama, pero no concuerdo con ninguna de ellas y me parece que no entiendo la segunda; pero como no son puntos de vista influyentes en la filosofía de la mente, y mis intentos de explicación apuntan a esta, no volveré a examinarlas en el libro.

Hay una versión más débil del dualismo denominada "dualismo de las propiedades", bastante difundi-

4 H. Stapp, *The Mindful Universe*, de próxima aparición.

5 La exposición clásica del idealismo se encontrará en G. Berkeley, *A Treatise Concerning the Principles of Human Knowledge*, edición establecida por J. Dancy, Oxford, Oxford University Press, 1998 [traducción española: *Tratado sobre los principios del conocimiento humano*, Madrid, Alianza, 1984].

do. La idea es la siguiente: en el mundo no hay dos tipos de sustancias, sino dos tipos de propiedades. La mayoría de estas, como el hecho de tener una carga eléctrica o una masa determinada, son propiedades físicas; pero algunas, como el hecho de sentir un dolor o pensar en Kansas City, son propiedades mentales. Aunque no están compuestos de dos clases diferentes de sustancias, los seres humanos exhiben la característica de que su cuerpo físico, y en particular su cerebro, tienen no sólo propiedades físicas sino también propiedades mentales.

El dualismo de las propiedades evita postular una sustancia mental independiente, pero hereda algunas de las dificultades del dualismo sustancial. ¿Cuáles son las relaciones entre lo mental y lo físico supuestas en esta doctrina? ¿Cómo pueden los sucesos físicos llegar a causar propiedades mentales? Por lo demás, hay un problema en particular que acosa a estos dualistas: cómo pueden las propiedades mentales, admitiendo que existan, actuar de manera causal para producir algo. ¿Cómo pueden mis estados conscientes, que según esta concepción ni siquiera son partes de una sustancia distinta, sino meros rasgos no físicos de mi cerebro, actuar y causar sucesos físicos en el mundo? En el capítulo 1 describí esta dificultad, cómo pueden los estados mentales actuar causalmente y producir efectos físicos, como el problema del "epifenomenalismo". De acuerdo con este, los estados mentales existen pero son epifenómenos. Se trata de simples compañeros de ruta; no tienen en realidad ningún efecto causal. Son como la espuma en la ola que llega a la orilla o los resplandores de luz que centellean en un lago: están ahí, pero no cumplen ningún papel causal significativo en el mundo físico. En rigor, son peores que la espuma y el resplandor, porque no podrían cumplir ningún

papel causal. El reto está en comprender cómo podrían cumplirlo en la determinación de los sucesos físicos, cuando ellos mismos no son físicos. Si suponemos, como al parecer debemos hacer, que el universo físico es causalmente cerrado, en el sentido de que nada exterior a él puede tener efecto en su interior; y si suponemos además, como al parecer debemos hacer, que la conciencia no forma parte del universo físico, habría que deducir que aquella no puede tener efecto alguno sobre este.

El dualismo de las propiedades no nos obliga a postular la existencia de una cosa que esté unida al cuerpo pero no sea realmente parte de él. Pero sí nos obliga a suponer que hay propiedades del cuerpo –presuntamente del cerebro– que no son propiedades físicas corrientes como el resto de nuestra constitución biológica. Y el inconveniente de esto es que no vemos de qué manera incorporar una descripción de esas propiedades a nuestra concepción global del universo y su modo de funcionamiento. En realidad no salimos de la postulación de entidades mentales por el hecho de llamarlas propiedades. Con ello seguimos postulando cosas mentales no materiales. No importa que digamos que mi dolor consciente es una propiedad mental de mi cerebro o que es un suceso dentro de este. De una u otra manera seguimos atrapados en las dificultades tradicionales del dualismo. Un filósofo antidualista caracterizó esos fenómenos mentales sobrantes como "rezagos nomológicos" ("nomológico" significa "con forma de ley"). El cerebro los produce a la manera de una ley, pero luego no hacen nada. Sencillamente se quedan ahí[6].

6 H. Feigl, "The 'Mental' and the 'Physical'", en H. Feigl, M. Scriven y G. Maxwell (comps.), *Concepts, Theories and the Mind-Body Problem*, Minneápolis, University of Minnesota Press, 1958, col. "Minnesota Studies in the Philosophy of Science", vol. 2.

Muchos filósofos, probablemente la mayoría, han abandonado el dualismo, pero la situación es curiosa porque numerosos dualistas consideran que los argumentos recién expuestos no parecen en modo alguno decisivos contra todas las formas de dualismo. Creo que un típico dualista de las propiedades diría: "De acuerdo, la mente no es una sustancia independiente, pero de todas maneras uno de los datos en bruto de la naturaleza es que las criaturas como nosotros tienen dolores, cosquillas y comezones, así como pensamientos y emociones, y estos no son físicos en un sentido corriente. Tampoco se los puede reducir a nada físico". Y en verdad, algunos dualistas hacen de tripas corazón y aceptan el epifenomenalismo.

Mi conjetura es que el dualismo, a pesar de estar pasado de moda, no desaparecerá. A decir verdad, esta doctrina –al menos en su versión de las propiedades– ha hecho en años recientes algo así como una reaparición, debida en parte al resurgimiento del interés en la conciencia. La intuición que lo impulsa es poderosa. Aquí la tenemos, en su presentación más simple: todos tenemos experiencias conscientes reales y sabemos que no son iguales a los objetos físicos que nos rodean. Podemos dar una forma más elaborada a esa intuición primigenia: el mundo esta hecho casi íntegramente de partículas físicas, y todo lo demás es en algún sentido una ilusión (como los colores y los sabores) o un rasgo superficial (como la solidez y la liquidez) que puede reducirse al comportamiento de aquellas partículas. En el nivel de la estructura molecular la mesa no es realmente sólida. Es, como dijo el físico Eddington, una

nube de moléculas. Sólo parece sólida desde nuestro punto de vista. Pero en el fondo el mundo físico está compuesto en su totalidad de microentidades, las partículas físicas. Hay una excepción, sin embargo. La conciencia no es sólo partículas. De hecho, no lo es en modo alguno. Sea lo que fuere, es algo que está "por encima" de las partículas. Creo que esta es la idea que le da fuerza al dualismo de las propiedades contemporáneo.

David Chalmers plantea este punto argumentando la imposibilidad lógica de que la trayectoria del universo físico sea diferente si la trayectoria de los hechos microfísicos es la misma[7]. Una vez que tenemos la microfísica todo lo demás puede deducirse. Pero esto no es válido para la conciencia. Podríamos imaginar que toda la trayectoria física del universo es exactamente la misma, menos la conciencia. Desde el punto de vista de la lógica es posible que esa trayectoria sea exactamente como es, pero sin conciencia.

Esas aparentes diferencias básicas entre lo mental y lo físico son el motor que impulsa el dualismo. Creo que este puede ser respondido y refutado, pero todavía no tenemos las herramientas para hacerlo. Me dedicaré a ello en el capítulo 4.

II. El giro hacia el materialismo

Los dualistas decían que hay dos clases de cosas o propiedades en el universo; con el fracaso del dualismo, es natural suponer que tal vez todo pertenezca a una sola clase. Esta perspectiva lleva el nombre poco sorprendente de "monismo" y se divide en dos: monismo mentalista y monismo materialista, llamados respectivamente "idealismo" y "materialismo". El idealismo dice que la totalidad del universo es mental o espiritual; sólo existen las "ideas" en el sentido técnico de la palabra, atribuido a todos los fenómenos mentales. Para algunas concepciones –la de Berkeley, por ejemplo–, además de las ideas están las mentes que las contienen. El idealismo tuvo durante varios siglos, sin exagerar, una prodigiosa influencia en la filosofía, pero que yo sepa ha estado muerto y enterrado desde hace muchas décadas para casi todos los filósofos cuyas opiniones respeto, por lo cual no me extenderé demasiado sobre él. Entre los más célebres idealistas cabe mencionar a Berkeley, Hegel, Bradley y Royce.

La familia más influyente de concepciones en la filosofía de la mente a lo largo del siglo XX y en estos comienzos del siglo XXI es alguna versión del materialismo. El materialismo es la noción de que la única realidad existente es la realidad material o física y, por consiguiente, si los estados mentales tienen existencia real, deben ser en cierto sentido reducibles a estados físicos de algún tipo, deben ser estados físicos. En ciertos aspectos el materialismo es la religión de nuestro tiempo, al menos entre la mayor parte de los profesionales expertos en el campo de la filosofía, la psicología, la ciencia cognitiva y otras disciplinas que estudian la mente. Como otras religiones más tradicionales, se lo acepta sin discusión y proporciona el marco dentro del cual es posible plantear, abordar y responder otras cuestiones. La historia del materialismo es fascinante, porque si bien los materialistas están convencidos, con una fe casi religiosa, de que su concepción debe ser correcta,

7 D. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford, Oxford University Press, 1996 [traducción española: *La mente consciente: en busca de una teoría fundamental*, Barcelona, Gedisa, 1999].

no parecen ser capaces de formular una versión que los satisfaga por completo y pueda ser aceptada en general por otros filósofos, aunque se trate de materialistas como ellos. A mi juicio, esto se debe a que tropiezan de manera constante contra el hecho de que diferentes versiones del materialismo parecen excluir algún rasgo mental esencial del universo, cuya existencia conocemos, cualesquiera sean nuestros compromisos filosóficos. Los rasgos habitualmente excluidos son la conciencia y la intencionalidad. El problema consiste en dar una descripción materialista totalmente satisfactoria de la mente que no termine por negar el hecho evidente de que todos tenemos en forma intrínseca estados conscientes y estados intencionales. En las próximas páginas voy a esbozar brevemente la historia del materialismo en el siglo XX, hasta el momento en que alcanzó por fin su formulación más sofisticada en la teoría computacional de la mente, según la cual el cerebro es una computadora y la mente es un programa informático. Por fuerza, el esbozo estará simplificado en exceso. Razones de espacio me obligan a destacar sólo los puntos culminantes, pero quiero que el lector los conozca y sepa cómo se relacionan entre sí. Hay una progresión natural que lleva desde el conductismo hasta la teoría computacional de la mente, y deseo exponerla.

III. La saga del materialismo: del conductismo a la inteligencia artificial fuerte

Conductismo

La primera forma influyente de materialismo en el siglo XX se denominó "conductismo". En su versión más cruda, esta doctrina dice que la mente es sólo el

comportamiento del cuerpo. Por encima de ese comportamiento no hay nada que sea constitutivo de lo mental. El conductismo se divide en dos tipos: "metodológico" y "lógico". Los examinaré en ese orden.

Conductismo metodológico

El conductismo metodológico fue un movimiento del ámbito de la psicología que intentó dar a esta disciplina un fundamento científico respetable y ponerla a la altura de las otras ciencias naturales. Con ese fin, insistía en que aquella sólo debía estudiar el comportamiento objetivamente observable. Las "leyes" que esa disciplina debía descubrir correlacionarían el estímulo de entrada al organismo [*input*] con la respuesta comportamental de salida [*output*]; por esa razón, la psicología conductista se denominó a veces psicología "del estímulo-respuesta". Los conductistas conquistaron tanta influencia que durante un tiempo lograron incluso modificar la definición de la psicología. Esta ya no era la "ciencia de la mente" sino la "ciencia del comportamiento humano". Esta corriente recibió el nombre de "conductismo metodológico" porque presentaba un método en psicología en vez de una proposición sustantiva acerca de la existencia o inexistencia de la mente. La verdadera objeción al dualismo, sostenían los conductistas metodológicos, no radica en su postulación de entidades no existentes, sino en su irrelevancia desde el punto de vista científico. Las proposiciones científicas deben ser verificables de manera objetiva, y las únicas proposiciones sobre la mente humana que cumplen esa condición son las referidas al comportamiento del hombre.

Los grandes nombres del conductismo metodológico son John B. Watson (1878-1958) y B. F. Skinner (1904-1990). A mi parecer, ninguno de ellos creía de hecho en la existencia de fenómenos mentales cualitativos internos, pero a efectos de constituir una psicología científica les era preciso insistir en el conductismo como un método y no como una doctrina ontológica específica. Acaso sea injusto caracterizar a Skinner como un conductista metodológico, porque en realidad planteaba objeciones a lo que denominaba "conductismo metodológico" y se consideraba un "conductista radical". No obstante, su influencia se ejerció sobre todo en el plano de la metodología; por eso, voy a seguir la exposición habitual de los libros de texto y lo caracterizaré como un conductista metodológico. Los únicos fenómenos psicológicos observables son los del comportamiento humano, de modo que el método apropiado para la psicología debe ser el estudio de ese comportamiento y no de misteriosas entidades mentales internas y espirituales. El conductismo metodológico fue, así, un proyecto de investigación en psicología y, sorpresivamente, gozó de influencia durante varias décadas.

Conductismo lógico

El conductismo lógico fue sobre todo un movimiento filosófico e hizo un planteo mucho más vigoroso que el conductismo metodológico. Los conductistas metodológicos decían que el dualismo cartesiano era irrelevante en términos científicos, mientras los conductistas lógicos sostenían que Descartes estaba equivocado por razones lógicas[8]. Un enunciado sobre el

8 Entre los conductistas lógicos de mayor celebridad se cuentan G. Ryle, del cual puede consultarse *The Concept of Mind*, *op. cit.*, y C. Hempel; de este, véase "The Logical Analysis of Psychology", en N. Block (comp.), *Readings in Philosophy of Psychology*, vol. 1, Cambridge (Mass.), Harvard University Press, 1980.

estado mental de una persona, decir por ejemplo que esta cree que va a llover o que siente un dolor en el codo, significa lo mismo que –o puede traducirse a– un conjunto de enunciados sobre su comportamiento real y posible. No es preciso que sea traducible en enunciados acerca de un comportamiento actualmente existente, pues la persona podría tener un dolor o una creencia sin manifestarlos de inmediato en una conducta; pero sí debe serlo en un conjunto de proposiciones hipotéticas sobre el comportamiento, lo que el agente haría o diría en tales o cuales circunstancias.

De conformidad con un análisis conductista típico, decir que Jones cree que va a llover es equivalente a plantear un número indefinido de proposiciones como las siguientes: si las ventanas de la casa de Jones están abiertas, este las cerrará; si las herramientas de jardinería quedaron a la intemperie, las guardará; si Jones sale a caminar, llevará un paraguas o se pondrá un impermeable, o ambas cosas, y así sucesivamente. La idea era que tener un estado mental sólo significaba estar dispuesto a exhibir ciertos tipos de comportamiento; el concepto de disposición, por su parte, debía analizarse en términos de proposiciones hipotéticas de la forma "si p, entonces q". Aplicadas al problema de los estados mentales, esas proposiciones asumirían la siguiente forma: "Si existen tales y cuales condiciones, resultará tal y cual comportamiento".

Fisicalismo y teoría de la identidad

Hacia mediados del siglo XX, las dificultades del conductismo habían provocado su debilitamiento generalizado y a la larga motivaron su rechazo. La doctrina no llevaba a ninguna parte como proyecto metodológico en psicología y a decir verdad era objeto de eficaces ataques, particularmente lanzados por el lingüista Noam Chomsky. Este afirmaba que la idea de que cuando estudiamos psicología estudiamos el comportamiento es tan poco inteligente como la idea de que cuando estudiamos física estudiamos lecturas de mediciones. Desde luego, utilizamos el comportamiento como prueba en psicología, así como usamos las lecturas de mediciones como prueba en física, pero es un error confundir la evidencia que tenemos acerca de un tema con el tema mismo. El tema de la psicología es la mente humana, y el comportamiento humano es la prueba de la existencia y los rasgos de esa mente, pero no es ella misma.

Las dificultades afrontadas por los conductistas lógicos eran aún más agudas. Nadie había propuesto una explicación siquiera remotamente plausible de cómo se podían traducir las proposiciones sobre la mente en proposiciones sobre el comportamiento. Había varias dificultades técnicas en lo concerniente a la manera de especificar los antecedentes de las hipótesis, y en especial cómo hacerlo sin caer en la circularidad. Dije antes que los conductistas descompondrían la creencia de Jones sobre la inminencia de la lluvia en conjuntos de proposiciones sobre su comportamiento para protegerse de esta. Pero el inconveniente radica en que sólo podemos comenzar a hacer esa reducción si suponemos que Jones desea mantenerse seco. Por lo tanto, el supuesto de que llevará un paraguas si cree que va a llover sólo es plausible si presumimos que no quiere mojarse. Pero si analizamos entonces la creencia en términos de deseo, parece haber una suerte de circularidad en la reducción. En realidad no hemos reducido la creencia al comportamiento; la redujimos al comportamiento más el deseo, con lo cual seguimos frente a un estado mental que es preciso analizar. Podrían hacerse observaciones análogas con respecto a la reducción del deseo. El argumento de que el deseo de Jones de estar seco consiste en cosas como su disposición a llevar un paraguas sólo parecerá remotamente plausible si suponemos que él prevé la proximidad de la lluvia.

Una segunda familia de dificultades se vinculaba con las relaciones causales entre estados mentales y comportamiento. Los conductistas lógicos habían argumentado que los estados mentales no consistían en otra cosa que comportamientos y disposiciones comportamentales, pero esta idea se opone a la intuición de sentido común de que hay relaciones causales entre nuestros estados mentales y nuestro comportamiento exterior. El dolor me lleva a gritar y tomar una aspirina; la creencia en que va a llover y el deseo de estar seco hacen que tome un paraguas, etc., y al parecer esta verdad evidente es negada por los conductistas, que no pueden explicar las relaciones causales entre la experiencia interna y el comportamiento externo porque niegan, en sustancia, la existencia de toda experiencia interna por añadidura al comportamiento externo.

La verdadera dificultad del conductismo, empero, es que su mero carácter poco plausible se convirtió en un estorbo cada vez más grande. Tenemos sin duda pensamientos, sentimientos, dolores, cosquillas y comezones, pero no parece razonable suponer que son idénticos a nuestro comportamiento, y ni siquiera a

nuestras disposiciones a adoptarlo. La sensación de dolor es una cosa, el comportamiento inducido por este, otra. Desde un punto de vista intuitivo el conductismo es tan poco convincente que a menudo los comentaristas poco afectos a él lo hacían objeto de sus burlas. Ya en la década de los veinte I. A. Richards señaló que para un conductista uno debe "fingir anestesia"[9]. Y los catedráticos universitarios tienen un repertorio habitual de malos chistes sobre esta doctrina. Un chiste típico: una pareja conductista acaba de hacer el amor y el hombre dice: "Fue fantástico para ti. ¿Cómo fue para mí?"

Hacia la década de los sesenta la completa inadmisibilidad del conductismo se había transformado en un impedimento, por lo cual los filósofos de inclinaciones materialistas lo reemplazaron poco a poco por una doctrina denominada "fisicalismo" y a veces "teoría de la identidad". Los fisicalistas decían que Descartes no estaba equivocado en el plano de la lógica –como habían sostenido los conductistas lógicos–, sino en el plano de los hechos. Podría haber sucedido que además de un cuerpo tuviéramos un alma, pero tal como resultaron las cosas en la naturaleza, lo que concebimos como mente es sólo un cerebro, y lo que imaginamos como estados mentales, por ejemplo la sensación de dolor o la impresión de tener cosquillas o una comezón, no son sino estados cerebrales, y tal vez del resto del sistema nervioso central. Esta postura recibió en ocasiones el

9 No puedo encontrar la fuente exacta de esta cita. Creo que es una adaptación de la caracterización de Ogden y Richards cuando señalan que Watson "simula una anestesia general". Véase C. K. Ogden e I. A. Richards, *The Meaning of Meaning* (1926), Londres, Harcourt Brace and Company, 1949, p. 23 [traducción española: *El significado del significado*, Barcelona, Paidós, 1984].

nombre de "tesis de la identidad", porque se afirmaba una identidad entre estados mentales y estados cerebrales. Los teóricos de la identidad procuraban insistir con afán en el contraste entre su concepción y el conductismo, visto como una tesis lógica sobre la definición de conceptos mentales. La tesis de la identidad, por su parte, era presuntamente una afirmación fáctica, no sobre el análisis de conceptos mentales, sino sobre el modo de existencia de los estados mentales. Los conductistas utilizaban el modelo de las identidades definicionales. Los dolores son disposiciones al comportamiento del mismo modo que los triángulos son figuras planas de tres lados. En cada caso es una cuestión de definición. Los teóricos de la identidad dijeron: no, el modelo no son las definiciones sino, antes bien, los descubrimientos empíricos de identidades en la ciencia. Hemos descubierto, de hecho, que un rayo es idéntico a una descarga eléctrica; hemos descubierto, de hecho, que el agua es idéntica a H_2O, y ahora descubrimos –un descubrimiento hecho día a día– que los estados mentales son en realidad idénticos a los estados cerebrales[10].

Objeciones a la teoría de la identidad

La teoría de la identidad recibió una serie de objeciones. Me parece útil distinguir entre las objeciones técnicas y las basadas en el sentido común. La primera

10 Se encontrarán tres exposiciones clásicas de la teoría de la identidad en U. T. Place, "Is Consciousness a Brain Process?", *British Journal of Psychology*, 47(1), 1956, pp. 44-50; J. J. C. Smart, "Sensations and Brain Processes", en D. Rosenthal (comp.), *The Nature of Mind*, Nueva York, Oxford University Press, 1991, pp. 169-176, y H. Feigl, "The 'Mental' and the 'Physical'", *op. cit.*

objeción técnica fue que la teoría parecía violar un principio lógico llamado "ley de Leibniz"[11]. Esta dice que si dos cosas cualesquiera son idénticas, deben tener todas sus propiedades en común. Por lo tanto, si pudiéramos mostrar que los estados mentales tienen propiedades imposibles de atribuirse a los estados cerebrales, y viceversa, al parecer refutaríamos la teoría de la identidad. Por lo demás, no parecía difícil proporcionar ejemplos al respecto. Así, puedo decir, pongamos por caso, que el estado cerebral correspondiente a mi pensamiento de que está lloviendo se encuentra tres centímetros dentro de mi oído izquierdo; pero, de acuerdo con los objetores, no tiene ningún sentido decir que mi pensamiento de que está lloviendo está tres centímetros dentro de mi oido izquierdo. Por otra parte, aun en el caso de los estados conscientes que tienen una localización, como el dolor, este puede situarse en un dedo del pie, pero el estado cerebral correspondiente no está en el dedo sino en el cerebro. Las propiedades del estado cerebral, entonces, no son iguales a las propiedades del estado mental. En consecuencia, el fisicalismo es falso.

Los teóricos de la identidad creían tener una respuesta simple a esas objeciones. Estas, decían, se apoyan en la ignorancia. Cuando sepamos más sobre el cerebro, llegaremos a juzgar perfectamente adecuada la atribución de localizaciones espaciales a los estados mentales y de las llamadas propiedades mentales a los estados cerebrales. Y con respecto a la localización del dolor en el dedo del pie, aquellos teóricos sostenían que

11 Esta objeción y las siguientes se analizan en J. J. C. Smart, "Sensations and brain processes", *op. cit.*

nuestro interés no estaba en el objeto putativo, el dolor, sino en la experiencia global de sentirlo. Y esa experiencia global abarca desde la estimulación de las terminaciones nerviosas periféricas del dedo hasta el propio cerebro. A mi parecer, los teóricos de la identidad lograron responder a esta objeción, pero había otras que eran más serias.

Una objeción de sentido común a la teoría de la identidad aducía que si esta era en efecto una identidad empírica, algo que podía descubrirse como un hecho, según la analogía del agua y el H_2O o el rayo y la descarga eléctrica, deberían existir dos tipos de propiedades para poder establecer con solidez ambos lados de la proposición de identidad[12]. De tal modo, así como el enunciado "el rayo es idéntico a una descarga eléctrica" debe identificar una y la misma cosa en términos de sus propiedades de rayo y de sus propiedades de descarga eléctrica, y el enunciado "el agua es idéntica a las moléculas de H_2O" debe identificar una y la misma cosa en términos de sus propiedades de agua y de sus propiedades de H_2O, la afirmación, por ejemplo, de que "el dolor es idéntico a cierto tipo de estado cerebral" tiene que identificar una y la misma cosa en términos de sus propiedades de dolor y de sus propiedades de estado cerebral. Pero si en la proposición de identidad hay dos conjuntos independientes de propiedades, es de presumir que nos quedan dos tipos diferentes de estas: las mentales y las físicas. En suma, parece como si, a fin de permitir la validez de la tesis de la identidad, tuvié-

12 Entre otros, esta objeción fue planteada por J. T. Stevenson, "Sensations and Brain Processes: A Reply to J. J. C. Smart", en C. V. Borst (comp.), *The Mind-Brain Identity Theory*, Nueva York, St. Martin's Press, 1970, pp. 87-92.

ramos que recaer en el dualismo de las propiedades. Si todos los estados mentales son estados cerebrales, hay dos clases de estos últimos, los que son mentales y los que no lo son. ¿Cuál es la diferencia? Los estados mentales tienen propiedades mentales. Los otros sólo tienen propiedades físicas. Y esa concepción se asemeja mucho al dualismo de las propiedades.

Este fue un problema decisivo para los teóricos de la identidad. Todo el sentido de la teoría radicaba en reivindicar el materialismo, mostrar que los estados mentales eran realmente idénticos a los estados materiales del cerebro: no eran otra cosa que estados materiales del cerebro y se los podía reducir a ellos. Pero si resulta que los estados mentales en cuestión tienen propiedades mentales irreductibles, el proyecto fracasa. Nos deja un elemento mental imposible de reducir. En mis investigaciones para este libro encontré como mínimo un filósofo que, aunque se consideraba un teórico de la identidad, parecía dispuesto a aceptar ese resultado, al menos como posibilidad[13]. Grover Maxwell da a su concepción el nombre de teoría de la identidad, pero dice: "el camino está totalmente abierto para especular que algunos sucesos mentales son simplemente nuestras alegrías, aflicciones, dolores, pensamientos, etc., en toda su riqueza cualitativa y mentalista" (p. 235). Esto es muy similar a la concepción que considero correcta, que explicaré en el capítulo 4. Pero no era una perspectiva típica entre los teóricos de la identidad.

13 G. Maxwell, "Unity of Consciousness and Mind-Brain Identity", en J. C. Eccles (comp.), *Mind and Brain: The Many Faceted Problems*, Washington, Paragon House, 1974, pp. 233-237.

La respuesta característica dada por estos a esa objeción fue menos convincente que su respuesta a las objeciones relacionadas con la ley de Leibniz[14]. Dijeron que los fenómenos en cuestión podían especificarse sin utilizar ningún predicado mental. Era posible hacerlo con un vocabulario coloquial neutral. En vez de decir: "Hay en mí una imagen residual entre amarilla y anaranjada", prefieren decir: "En mí sucede algo semejante a lo que ocurre cuando veo una naranja". Supuestamente, esa reformulación de la identificación de los estados mentales en un vocabulario "coloquial neutral" respondía a la objeción, porque nos permitía especificar el elemento mental en un léxico neutro y no mental: en mí sucede una cosa que puede especificarse de una manera neutral entre el dualismo y el materialismo, pero resulta justamente que la cosa es un proceso cerebral. Así, podemos dar especificidad al rasgo mental, pero de un modo compatible con el materialismo.

Creo que esta respuesta es fallida. El argumento de que podemos hablar de los fenómenos mentales sin utilizar un vocabulario mental no modifica el hecho de que esos fenómenos siguen teniendo propiedades mentales. Mi imagen residual entre amarilla y anaranjada sigue siendo cualitativa y subjetiva al margen de que decidamos mencionar u omitir esas características. Si uno quisiera negarse a hablar de aviones, le bastaría con decir: "algún bien perteneciente a United Airlines". Pero eso no suprime la existencia de los aviones. Para expresarlo de manera sucinta, la referencia a un fenó-

14 Esta objeción se discutió en el artículo original de Smart, y también en J. J. C. Smart, "Further Remarks on Sensations and Brain Processes", en V. Borst (comp.), *The Mind-Brain Identity Theory*, *op. cit.*, pp. 93-94.

meno que es intrínsecamente cualitativo y subjetivo en un vocabulario que no revela esos rasgos no elimina estos últimos. En resumidas cuentas, los teóricos de la identidad pretendían negar la existencia de tales rasgos, pero eso exige otro argumento[15].

Una objeción levemente más técnica que en realidad preocupó a los teóricos de la identidad y a la larga los obligó a modificar sus concepciones fue la acusación de "chovinismo neuronal"[16]. Si la tesis de esos teóricos era que todo dolor es idéntico a cierto tipo de estimulación neuronal, y toda creencia es idéntica a cierto tipo de estado cerebral, parece deducirse que un ser sin neuronas o al menos sin la clase apropiada de ellas no podría tener dolores y creencias. Pero ¿por qué los animales con estructuras cerebrales diferentes de la nuestra no pueden tener estados mentales? Y en rigor, ¿por qué no podríamos construir una máquina que no tuviera absolutamente ninguna neurona, pero sí estados mentales? Esta objeción provocó un cambio importante en la teoría de la identidad; se pasó de lo que llegó a llamarse "teoría de la identidad tipo-tipo" a la "teoría de la identidad caso-caso". Para explicar esta distinción es preciso decir algunas palabras sobre la diferencia entre tipo [*type*] y caso [*token*]. Si escribo la palabra "perro" tres veces: "perro perro perro", ¿he escrito una palabra o tres? Bueno, he escrito tres ejemplos o casos de un tipo de palabra. De modo que necesitamos una

15 J. R. Searle, *The Rediscovery of the Mind*, *op. cit.*

16 N. Block, "Troubles with Functionalism"en C. Wade Savage (comp.), *Perception and Cognition: Issues in the Foundations of Psychology*, vol. 9, Minneápolis, University of Minnesota Press, 1978, col. "Minnesota Studies in the Philosophy of Science", pp. 261-325, reeditado en N. Block (comp.), *Readings in Philosophy of Psychology*, *op. cit.*, pp. 268-305.

distinción entre tipos, que son entidades generales abstractas, y casos, que son objetos y sucesos particulares y concretos. El caso de un tipo es una ejemplificación particular concreta de ese tipo general abstracto.

A través de esa distinción podemos ver por qué los teóricos de la identidad sintieron la necesidad de pasar de una teoría tipo-tipo a una teoría caso-caso. La teoría de la identidad tipo-tipo dice: "Todo tipo de estado mental es idéntico a algún tipo de estado físico". Esta afirmación es a todas luces un poco chapucera, porque la identidad en cuestión es la existente entre casos reales concretos y no entre tipos universales abstractos. Lo que esos teóricos quieren decir es: para cada tipo de estado mental hay algún tipo de estado cerebral tal que cada caso del tipo mental es un caso del tipo cerebral. Los teóricos de la identidad de casos decían simplemente: para cada caso de un tipo determinado de estado mental hay algún caso de algún tipo de estado físico idéntico a ese caso de estado mental. En síntesis, no exigían, digamos, que todos los casos de dolores tuvieran que ejemplificar exactamente el mismo tipo de estado cerebral. Podía tratarse de casos de diferentes tipos de estados cerebrales, aun cuando todos fueran casos del mismo tipo mental, el dolor. Por esa razón se les dio el nombre de teóricos de la identidad "caso-caso", en contraste con los teóricos de la identidad "tipo-tipo". La identidad entre casos parece mucho más plausible que la identidad entre tipos. Supongamos que tanto usted como yo creemos que Denver es la capital de Colorado. Parece innecesario suponer que, para tener la misma creencia, usted y yo debemos encontrarnos exactamente en el mismo tipo de estado neurobiológico. El estado neurobiológico por el cual creo que Denver es la capital de Colorado podría localizarse en

un punto determinado de mi cerebro, y el suyo podría situarse en otro punto, sin que se tratara de creencias diferentes.

Desafortunadamente, los teóricos de la identidad propusieron a menudo ejemplos bastante pobres. Uno de los favoritos consistía en decir que los dolores son idénticos a las estimulaciones de las fibras C. La idea era que, de acuerdo con los teóricos de la identidad de tipos, todo dolor es idéntico a alguna estimulación de las fibras C; según los teóricos de la identidad de casos, tal dolor en particular podía ser idéntico a tal estimulación en particular de las fibras C, pero otro dolor podía ser idéntico a algún otro estado del cerebro o de una máquina. Es una lástima que todo esto sea neurofisiología bastante mala. Una fibra C es un tipo de axón, y es cierto que algunos tipos de señales de dolor, no todos, son transmitidos por esas fibras al cerebro. Pero desde un punto de vista neurofisiológico sería ridículo creer que los dolores no consisten en nada más que la estimulación de nuestras fibras C. Estas sólo son parte de un complejo mecanismo del dolor en el cerebro y el sistema nervioso. Sea como fuere, esa fue la clase de ejemplos presentada por los teóricos de la identidad, y buena parte del debate se centró en determinar si obtendríamos esas identidades de tipos o sólo cabía esperar identidades de casos. A la postre, los teóricos de la identidad de casos han ejercido mayor influencia que los teóricos de la identidad de tipos.

Pero ahora nos enfrentamos a una cuestión interesante. ¿Qué tienen en común todos esos casos para ser casos del mismo tipo de estado mental? Si usted y yo creemos que Denver es la capital de Colorado, ¿qué es exactamente lo que compartimos, si no hay otra cosa que nuestros estados cerebrales y estos son de diferente

tipo? Adviértase que las dos respuestas que tradicionalmente se darían a esta pregunta, la dualista y la de la identidad tipo-tipo, son inaceptables para el fisicalista de casos. Este no puede decir que su factor en común son las mismas propiedades irreductiblemente mentales, porque todo su propósito era eliminarlas o deshacerse de ellas. Tampoco puede aducir que se trata del mismo tipo de estado cerebral, porque la razón para pasar de la teoría de la identidad de tipos a la teoría de la identidad de casos fue no tener que decir que cada caso de un tipo de estado mental determinado es idéntico a un caso de un tipo de estado cerebral determinado.

Funcionalismo

En este punto los materialistas dieron un paso que fue crucial para el ulterior filosofar sobre la mente. Dijeron: si los casos de estados cerebrales son estados mentales es porque tienen cierto tipo de función en el comportamiento general del organismo. No es una sorpresa que esta doctrina se denominara "funcionalismo", y al desplegarse derivó en concepciones como la siguiente[17]: decir que Jones cree que está lloviendo es decir que en él se desenvuelve cierto suceso, estado o proceso causado por determinada clase de estímulos

17 Entre los primeros partidarios del funcionalismo se cuentan H. Putnam, D. Lewis y D. Armstrong. Véanse H. Putnam, "The Nature of Mental States", en N. Block (comp.), *Readings in Philosophy of Psychology*, *op. cit.*, pp. 223-231 [traducción española: *La naturaleza de los estados mentales*, México, Instituto de Investigaciones Filosóficas de la UNAM, 1981]; D. Lewis, "Psychophysical and Theoretical Identifications" y "Mad Pain and Martian Pain", en *ibid.*, pp. 207-215 y 216-222 respectivamente, y D. Armstrong, *A Materialist Theory of Mind*, Londres, Routledge, 1993.

externos, por ejemplo, la percepción de la lluvia; y este fenómeno, en conjunción con algunos otros factores como su deseo de mantenerse seco, generarán en nuestro hombre un comportamiento determinado, el de tomar un paraguas. En síntesis, los estados mentales se definen como estados con ciertas funciones, y el concepto de función se explica en términos de relaciones causales con estímulos externos, otros estados mentales y el comportamiento externo. Podríamos formular así este desarrollo: la percepción de la lluvia causa en Jones la creencia de que llueve. Esa creencia y el deseo de no mojarse causan el comportamiento consistente en tomar el paraguas. ¿Qué es, entonces, una creencia? Todo lo que se inscribe en esa clase de relaciones causales. En este punto los teóricos de la identidad introdujeron un hermoso dispositivo técnico para capturar precisamente ese rasgo de su teoría. El dispositivo recibió el nombre de "cláusula de Ramsey" por su inventor, el filósofo británico Frank Ramsey. En la conjunción anterior de oraciones simplemente eliminamos "en Jones la creencia de que llueve" y la reemplazamos por *x*. Luego anteponemos a toda la frase un cuantificador existencia que dice "hay un *x* tal que". De modo que ahora reza así: "hay un *x* tal que la percepción de la lluvia causa *x*, y *x* junto con el deseo de no mojarse causan el comportamiento consistente en tomar un paraguas". Por eso, ¿qué es realmente una creencia? Es cualquier cosa, cualquier *x* que se encuentra en esas relaciones causales (y muchas otras semejantes). Los estados mentales como las creencias no se definen por ninguna característica intrínseca sino por sus relaciones causales, y estas constituyen su función. Las creencias, por ejemplo, son causadas por percepciones, y junto con los deseos causan acciones. Las relaciones

causales son el único contenido del hecho de tener una creencia.

¿Y qué pasa con la referencia restante a los deseos y las percepciones? También ellos se analizarán desde un punto de vista funcional. Así como hay un *x* que es la creencia, definida por sus relaciones causales, hay un *y* que es el deseo y un *z* que es una percepción, y uno y otra también se definen por sus relaciones causales.

La descripción funcionalista hizo frente entonces a varias de las objeciones al conductismo. Una de ellas era su aparente circularidad en el uso de los deseos para explicar las creencias y de estas para explicar aquellos. El funcionalista da una rápida respuesta a esta objeción, si se analizan las creencias y los deseos de manera simultánea, en términos de sus relaciones causales. También respondemos de inmediato la objeción de que el conductismo excluyó las relaciones causales entre estados mentales y comportamiento externo, porque hemos definido en parte los primeros desde la perspectiva de su capacidad de causar un comportamiento externo. Por lo demás, un atractivo adicional de la explicación funcionalista de los estados mentales es que parecía asimilar el reino mental a un reino muy conocido de entidades funcionales humanas. Así, si preguntamos: ¿qué es un carburador, un termostato, un reloj?, todas estas preguntas se responden causalmente describiendo las funciones causales de carburadores, termostatos y relojes. Ninguna de estas cosas se define por su estructura física. Un reloj, por ejemplo, puede estar compuesto de engranajes y ruedas, de dos ampollas de vidrio unidas por el cuello y con arena en su interior, de osciladores de cuarzo o de muchos otros materiales físicos, pero su rasgo definitorio es que se trata de un mecanismo físico que nos permite saber la hora. Po-

drían hacerse observaciones análogas sobre carburadores y termostatos. Los estados mentales son semejantes a los carburadores, los termostatos y los relojes. No se definen por su estructura física ni por una esencia mental cartesiana; antes bien, las relaciones causales son su elemento definitorio. Una creencia es cualquier entidad que, situada en ciertas relaciones con los estímulos entrantes y otros estados mentales, es la causa de un comportamiento externo.

El impulso subyacente del funcionalismo era responder la siguiente pregunta: ¿por qué atribuimos estados mentales a las personas? Y la respuesta era: decimos que tienen cosas tales como creencias y deseos porque queremos explicar su comportamiento. El funcionalismo parece haber aprehendido todas esas intuiciones.

Como es comprensible, los funcionalistas querían saber cuál era la naturaleza de los estados cerebrales y mentales internos que les permitía causar un comportamiento. ¿Cuál era la diferencia entre los estados mentales y otros tipos de estados cerebrales? Una respuesta consistía en decir que esa pregunta no es adecuada en modo alguno para la filosofía; habría que plantearla a psicólogos y neurobiólogos. Podemos tratar el cerebro como una mera "caja negra" que produce comportamientos en respuesta a estímulos, y no es necesario que, como filósofos, nos preocupemos por el mecanismo existente en su interior. En ocasiones, esta concepción recibía el nombre de "funcionalismo de la caja negra".

Pero el funcionalismo de la caja negra es intelectualmente insatisfactorio porque no da respuestas a nuestra natural curiosidad intelectual. Queremos saber, en realidad, cómo funciona el sistema.

Funcionalismo computacional (= inteligencia artificial fuerte)

En este punto se produjo uno de los más fascinantes desarrollos de toda la historia de la filosofía de la mente en el siglo XX. Para muchos de quienes participaron en él (aunque no para mí), ese desarrollo fue no sólo fascinante sino una solución, por fin, a problemas que habían asediado a los filósofos durante más de dos mil años. La idea se basaba en una convergencia de trabajos en filosofía, psicología cognitiva, lingüística, informática e inteligencia artificial. Al parecer, teníamos la respuesta a la cuestión que enfrentábamos, cómo funciona el sistema: el cerebro es una computadora digital y lo que llamamos "mente" es un programa o conjunto de programas informáticos digitales. Habíamos hecho el más grande avance en la historia de la filosofía de la mente: los estados mentales son estados computacionales del cerebro. Este es una computadora y la mente es un programa o conjunto de programas. Una enorme cantidad de libros de texto se fundaron en este principio: la mente es al cerebro lo que el programa es al *hardware*[18].

$$\frac{\text{Mente}}{\text{Cerebro}} = \frac{\text{Programa}}{\text{Hardware}}$$

18 P. Johnson-Laird, *The Computer and the Mind*, Cambridge (Mass.), Harvard University Press, 1988 [traducción española: *El ordenador y la mente*, Barcelona, Paidós, 1990], y *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness*, Cambridge (Mass.), Harvard University Press, 1983.

Esta perspectiva se denomina a veces "funcionalismo computacional", aunque yo también la bauticé "inteligencia artificial fuerte" para distinguirla de la inteligencia artificial débil, que, en contraste con el propósito de crear una mente, aspira a estudiarla mediante simulaciones por computadora. Según el punto de vista de la inteligencia artificial fuerte, con la programación adecuada la computadora digital no simula tener una mente: la tiene literalmente.

Con la aparición del modelo computacional de la mente creímos haber encontrado por fin la solución a los problemas que habían inquietado a Descartes e incluso a los primeros filósofos griegos, dos mil quinientos años atrás. En especial, teníamos en apariencia una solución perfecta para el tradicional problema mente-cuerpo. La relación entre una y otro parecía misteriosa; en cambio, la existente entre el programa y el *hardware* informático, la relación del *software* con su implementación física, no lo es en lo más mínimo. Se la entiende en todos los departamentos de informática del mundo, y ese conocimiento se utiliza de manera rutinaria y cotidiana para programar computadoras.

IV. La computación y los procesos mentales

Hasta aquí he criticado las concepciones materialistas según su orden de aparición. Pero ahora voy a exponer la teoría computacional de la mente y reservaré las críticas dirigidas a ella y otras versiones del funcionalismo hasta el próximo capítulo. Antes de explicar en detalle las supuestas soluciones aportadas por esa teoría computacional a nuestros problemas, quiero introducir varias nociones cruciales. Estas son importantes por su pertinencia no sólo para la filosofía contemporánea

sino, a decir verdad, para la vida intelectual en general. Las nociones que espero explicar con brevedad son las de algoritmo, máquina de Turing, tesis de Church, teorema de Turing, prueba de Turing, niveles de descripción, realizabilidad múltiple y descomposición recursiva. Estos conceptos son el núcleo de lo que hasta hace poco fue, y en algunos ámbitos todavía es, la visión más influyente de la naturaleza de la mente en la ciencia cognitiva y disciplinas conexas. Por otra parte, varias de estas ideas son tan importantes que es esencial para la educación general del lector, al margen de la filosofía, familiarizarse plenamente con esos conceptos.

Algoritmos. Un algoritmo es un método para resolver un problema a través de una serie precisa de pasos. Los pasos deben ser finitos en número y su correcta realización garantiza la solución del problema. Por ese motivo, los algoritmos también reciben el nombre de "procedimientos eficaces". Buenos ejemplos son los métodos utilizados para resolver problemas en aritmética, como la suma y la resta. Si seguimos los pasos con exactitud, llegaremos a la solución correcta.

Máquinas de Turing. Una máquina de Turing es un dispositivo que realiza cálculos empleando sólo dos tipos de símbolos. En general se supone que estos son ceros y unos, pero cualquier símbolo podría servir. La concepción de esta máquina se debe a Alan Turing, el gran lógico y matemático británico. La característica más llamativa del dispositivo es su simplicidad: tiene una cinta sin fin en la cual se escriben los símbolos y una cabeza que los lee. Esta cabeza se mueve hacia la izquierda o hacia la derecha y puede borrar un cero e imprimir un uno o borrar un uno e imprimir un cero.

Hace todas estas cosas de conformidad con un programa, que consiste en un conjunto de reglas. Las reglas siempre tienen la misma forma; en la condición C, ejecute el acto A: C → A. Una regla podría tener, por ejemplo, la siguiente forma: si está examinando un cero, reemplácelo por un uno y muévase un espacio a la izquierda.

La máquina de Turing no es una máquina en el sentido habitual. No es posible comprarla en una tienda. Es un concepto matemático abstracto. Por ejemplo, tiene una cinta sin fin y, por ende, una capacidad infinita de almacenamiento. Ninguna máquina real tiene esa característica. Las máquinas de verdad se descomponen, se oxidan o se les cae cerveza encima. Las máquinas de Turing no tienen ninguno de esos defectos porque son puramente abstractas. Sin embargo, aunque su concepto es el concepto de algo formal y abstracto, a los efectos prácticos el tipo de computadora que compramos en una tienda es una máquina de Turing. Las computadoras comerciales corrientes implementan algoritmos mediante la manipulación de dos clases de símbolos. La electrónica contemporánea es tan sofisticada que la computadora de nuestros días puede llevar a cabo esas operaciones simbólicas a una velocidad de millones por segundo.

Tesis de Church. Debida en su origen a Alonzo Church (aunque Turing llegó de manera independiente a ella, por lo cual a veces se la llama tesis de Church-Turing), esta tesis sostiene que cualquier problema que tenga una solución algorítmica puede resolverse por medio de una máquina de Turing. O, según otra manera de decirlo: cualquier algoritmo puede llevarse a cabo en una máquina de Turing. La idea de una máquina que

sólo utilice símbolos binarios, ceros y unos, es suficiente para realizar absolutamente cualquier algoritmo. Esta tesis es muy importante, porque dice en términos matemáticos que cualquier problema computable puede computarse en una de esas máquinas. Cualquier función computable es computable *à la* Turing.

Las máquinas de Turing pueden presentarse en muchos tipos, estados y variedades diferentes. En mi automóvil hay computadoras especializadas para detectar el promedio de consumo de combustible, por ejemplo. Pero además de la idea de estas computadoras con finalidades especiales, o máquinas de Turing, está la idea de una computadora multipropósito, un dispositivo capaz de ejecutar cualquier programa. Y Alan Turing, en un importante resultado matemático conocido como teorema de Turing, demostró que hay una máquina universal de Turing que puede simular el comportamiento de cualquier otra de tales máquinas. Más precisamente, demostró que hay una máquina universal de Turing, UTM [*Universal Turing Machine*], tal que, dada cualquier máquina de Turing que ejecute un programa específico, TP, la UTM puede ejecutarlo.

La fascinación despertada por esas ideas se explica por la siguiente conjetura: ¿qué pasa si suponemos que el cerebro humano es una máquina universal de Turing? No puedo describir la excitación generada por esta idea, que nos daba por fin no sólo una solución a los problemas filosóficos que nos atormentaban, sino también un programa de investigación. Podemos estudiar la mente, averiguar cómo funciona realmente, si descubrimos qué programas se implementan en el cerebro. Una característica de enorme atractivo de ese programa de investigación es que en realidad no tenemos que saber cómo funciona el cerebro en cuanto sistema

físico para hacer una ciencia cabal y estricta de la mente. Las especificidades del cerebro son en verdad irrelevantes para la mente, porque cualquier otro sistema físico serviría, con tal de que fuera suficientemente estable y rico para contener los programas. De acuerdo con este punto de vista, los pormenores neurobiológicos del funcionamiento cerebral no tienen importancia para la mente. Por una especie de accidente evolutivo, la casualidad quiso simplemente que tuviéramos neuronas, pero cualquier sistema de *hardware* lo bastante complejo serviría tan bien como lo que tenemos dentro del cráneo. Para llegar a una descripción científica realmente adecuada de la mente, no hace falta más que descubrir los programas de la máquina de Turing que todos utilizamos en nuestros procesos de cognición.

El test de Turing. Sin embargo, necesitamos una prueba. Necesitamos una prueba que nos diga cuándo una máquina se comporta de manera auténticamente inteligente y cuándo no lo hace. Su invención también correspondió a Alan Turing, y por eso se la denomina test de Turing. Hay distintas versiones, pero la idea básica es la siguiente: para eludir los grandes debates acerca del problema de la existencia de otras mentes y del pensamiento y la inteligencia presuntos de la máquina, basta con preguntarse si esta puede desenvolverse de tal manera que un experto sea incapaz de distinguir su desempeño de un desempeño humano. Si la máquina contesta preguntas formuladas en chino con tanta aptitud como un hablante nativo de esa lengua, de modo tal que otros hablantes nativos sean incapaces de ver la diferencia entre aquella y uno cualquiera de ellos, deberemos decir que la máquina entiende el chino. Como el lector habrá advertido, el test de Turing es expresión

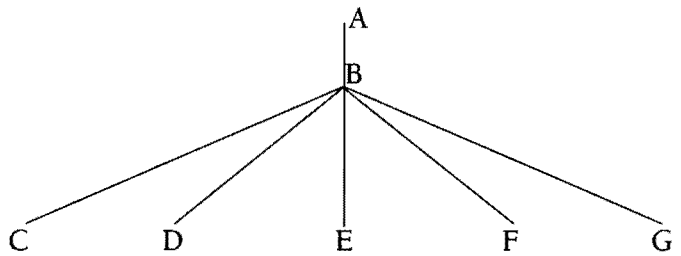
de una especie de conductismo. Dice que la prueba comportamental es concluyente acerca de la presencia de estados mentales.

Niveles de descripción. Cualquier sistema complejo puede describirse de diferentes maneras. Así, por ejemplo, el motor de un automóvil puede caracterizarse en términos de su estructura molecular, de su forma física general, de sus partes componentes, etc. Es tentador presentar esta variabilidad de posibilidades descriptivas según la metáfora de los "niveles", terminología que ha ganado una aceptación generalizada. Concebimos el micronivel de las moléculas como un nivel de descripción más bajo que el de la estructura física general o los componentes materiales, que son niveles descriptivos más elevados. Casi todo el interés de esta distinción estriba en su contundente validez para las computadoras. En un nivel inferior de descripción, tu computadora y la mía pueden ser muy diferentes. La tuya quizá tenga un tipo de procesador distinto del mío, por ejemplo. Pero en un nivel superior de descripción acaso implementen exactamente el mismo algoritmo y ejecuten el mismo programa.

Realizabilidad múltiple. La noción de diferentes niveles de descripción ya contiene de manera implícita otra idea decisiva para la teoría computacional de la mente, la de la realizabilidad múltiple. El argumento es que una característica de nivel más elevado, como el hecho de ser el programa Word o un carburador, puede realizarse materialmente en diferentes sistemas; de tal modo, es posible decir que una y la misma característica de superior nivel puede ser realizable de variadas maneras en distintos soportes de menor nivel. La

realizabilidad múltiple parece ser una característica natural de las teorías de la identidad de casos. Los distintos casos de distintos tipos del nivel inferior pueden ser diferentes formas de realización de algún rasgo mental común de nivel superior. Así como el mismo programa informático puede ejecutarse en diferentes clases de *hardware* y por eso es realizable de manera múltiple, el mismo estado mental, por ejemplo la creencia de que va a llover, podría implementarse en diversas clases de soporte y, con ello, ser también realizable de múltiples formas.

El siguiente diagrama ilustra la distinción entre niveles de descripción y la realizabilidad múltiple del nivel superior en niveles inferiores:



Un único sistema, representado por la línea AB, puede realizarse en diferentes sistemas de nivel inferior, representados por las líneas BC, BD, BE, BF y BG.

Descomposición recursiva. Otra idea importante, ya implícita en lo que he dicho, es que los grandes problemas complejos pueden descomponerse en pequeños problemas simples, susceptibles a su vez de descomponerse en problemas aún más simples, hasta alcanzar el nivel de simplicidad máxima. La multiplicación con varios dígitos, por ejemplo 28 x 71, puede parecernos una operación compleja, pero la belleza de la idea de la

máquina de Turing es que, en el fondo, esos problemas se descomponen hasta ser sencillas maniobras con ceros y unos. Imprimimos un uno, borramos un cero, nos movemos un espacio a la izquierda o a la derecha. Eso es todo lo que la máquina necesita saber hacer a fin de realizar no sólo aritmética sino los algoritmos más increíblemente complejos para otros tipos de tareas. Las tareas complejas pueden analizarse (descomponerse) en tareas simples mediante la aplicación repetida (recursiva) de los mismos procedimientos, hasta que sólo quedan sencillas operaciones binarias con dos símbolos, los ceros y los unos. En los primeros y embriagadores días, algunas personas llegaron incluso a decir que el hecho de que las neuronas hicieran una de dos cosas, activarse o no activarse, era una indicación de que el cerebro era un sistema binario, como cualquier otra computadora digital. La idea de la descomposición recursiva también parecía darnos, entonces, una pista importante para entender la inteligencia humana. Las tareas humanas inteligentes y complejas pueden descomponerse recursivamente en tareas simples, y por eso somos tan inteligentes.

El conjunto de ideas que acabo de exponer contiene las herramientas necesarias para enunciar la teoría de la mente más influyente y pujante de las últimas décadas del siglo XX. El cerebro es una computadora digital; con toda probabilidad, una máquina universal de Turing. Como tal, lleva a cabo algoritmos mediante la implementación de programas, y lo que llamamos mente es uno de esos programas o conjunto de programas. Para comprender las capacidades cognitivas humanas sólo es necesario descubrir los programas que los seres humanos ejecutan efectivamente cuando activan capacidades cognitivas como la percepción, la

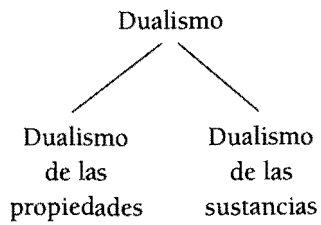
memoria, etc. Como el nivel mental de descripción es un nivel de programa, no nos es preciso entender los detalles del funcionamiento cerebral para entender la cognición humana. En rigor, al ser el nivel de descripción más elevado que el de las estructuras neuronales, no estamos obligados a adoptar ninguna teoría de la identidad tipo-tipo de la mente. Antes bien, los estados mentales son realizables de manera múltiple en diferentes clases de estructuras físicas, y si bien un azar los llevó a ejecutarse en el cerebro, podrían haberlo hecho con igual eficacia en una gama indefinida de soportes computacionales. La ejecución en cualquier soporte servirá para la mente humana, con la única condición de que sea lo bastante estable y rica para contener los programas. Como somos máquinas de Turing, seremos capaces de entender la cognición si reducimos las operaciones complejas a las operaciones más simples de todas, la manipulación de ceros y unos. Por lo demás, contamos con una prueba que nos permitirá constatar la reproducción efectiva de la cognición humana, el test de Turing. Este nos da una demostración concluyente de la presencia de capacidades cognitivas. Para averiguar si hemos inventado concretamente una máquina inteligente, sólo necesitamos aplicar el test de Turing. Y ahora tenemos un proyecto de investigación; se trata, en efecto, del proyecto de investigación de la ciencia cognitiva.

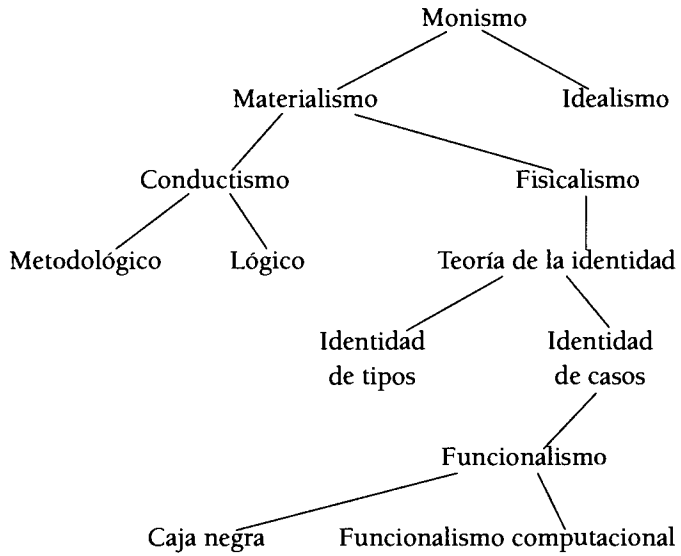
Tratamos de descubrir los programas que se ejecutan en el cerebro a través del diseño de programas para nuestras máquinas comerciales que pasen la prueba de Turing, y luego pedimos a los psicólogos que lleven a cabo experimentos con seres humanos a fin de ver si siguen el mismo programa que hemos incorporado a

nuestra computadora. Por ejemplo, en un famoso experimento relacionado con la recordación de números, los tiempos de reacción de los sujetos parecían variar de la misma manera que el tiempo de procesamiento de una computadora. Muchos especialistas en ciencia cognitiva consideraron este resultado como una prueba de que los seres humanos utilizaban los procedimientos algorítmicos de los ordenadores.

Tal fue el atractivo de la teoría computacional de la mente en los primeros días de la ciencia cognitiva. Si no he logrado que el lector la vea de ese modo, significa que mi exposición no ha sido buena; en su época, para muchos resultó enormemente interesante. La teoría generó millares de proyectos de investigación y acumuló un número parecido de subsidios para llevarlos a cabo. Pero, ay, es una teoría errónea más allá de toda esperanza. Así lo creí en esos momentos, y desde entonces nada me motivó a cambiar de opinión. En el próximo capítulo explicaré por qué está equivocada. Por ahora, quiero que el lector sepa apreciar su atractivo.

Con ciertas vacilaciones (porque es una simplificación excesiva), presento un cuadro que muestra las relaciones entre las teorías expuestas hasta aquí.





V. Otras versiones del materialismo

Una de las características interesantes del materialismo es que sus distintos representantes han adoptado virtualmente todas las posiciones materialistas concebibles. Para completar el relato del materialismo moderno, quiero mencionar otras dos versiones: el materialismo eliminativo, la idea de que los estados mentales no existen en absoluto, y el monismo anómalo, una idea de Donald Davidson que es una versión de la teoría de la identidad de casos.

El materialismo eliminativo sostiene lo siguiente[19]. ¿Por qué decimos que las personas tienen creencias, deseos y otras clases de estados mentales? Lo decimos porque queremos explicar su comportamiento. Por lo tanto, nuestra postulación de creencias, deseos, etc., es la postulación de un tipo de entidad teórica, así como en física lo son los postulados sobre los electrones o la fuerza electromagnética. Lo característico de esas postulaciones es que basta con demostrar la falsedad de la teoría para establecer la inexistencia de la entidad. La hoy obsoleta teoría del flogisto, según la cual la combustión de un objeto consistía en la liberación de una sustancia llamada "flogisto", ha sido refutada, y en virtud de su refutación ya no creemos en la existencia de dicha sustancia. ¿Cuál es, entonces, la teoría que postula creencias, deseos, etc.? Bueno, es la psicología del sentido común o de las abuelas, que en la literatura suele denominarse "psicología popular". Ahora bien, continúa el relato, es casi seguro que la psicología popular debe demostrar ser una teoría inadecuada y, en rigor, falsa. ¿Por qué? Por un lado, porque el progreso científico siempre ha refutado las teorías populares. Además, la psicología popular no conduce a ninguna parte como programa de investigación. Nuestras teorías populares de la racionalidad, por ejemplo, no mejoran demasiado la teoría de Aristóteles. Pero si la teoría que postula creencias, deseos, etc., es falsa, debe deducirse que esas entidades no existen. De modo que el materialismo eliminativo se limita a ser una versión del

19 En su origen, el eliminativismo fue propuesto por R. Rorty y P. Feyerabend. Uno de sus partidarios recientes es Paul Churchland. Véanse P. Feyerabend, "Mental Events and the Brain", *Journal of Philosophy*, 60, 1963, pp. 295-296; R. Rorty, "Mind-Body Identity, Privacy and Categories", en D. Rosenthal (comp.), *Materialism, and the Mind-Body Problem*, Englewood Cliffs (NJ), Prentice Hall, 1971, pp. 174-199, y P. M. Churchland, "Eliminative Materialism and the Propositional Attitudes", en D. Rosenthal (comp.), *The Nature of Mind*, *op. cit.*, pp. 601-612 [traducción española: "El materialismo eliminativo y las actitudes proposicionales", en Eduardo Rabossi (comp.), *Filosofía de la mente y conciencia cognitiva*, Barcelona, Paidós, 1995, pp. 43-68].

materialismo que suprime por completo los estados mentales. Se demuestra que estos son ilusiones: tienen el mismo carácter ilusorio que la puesta del sol y el flogisto.

Un argumento conexo contra las entidades de la psicología popular aludió a la falta de reducciones tipo-tipo de las nociones de esa psicología a fenómenos neurobiológicos. Es muy improbable que una neurociencia madura haga mucho uso de nociones como la creencia y el deseo, porque no son compatibles con las categorías de la neurobiología. Ante la ausencia de una reducción tipo-tipo de las creencias y los deseos, parece razonable suponer que esas entidades no existen.

El monismo anómalo es una concepción expuesta por Donald Davidson[20], en cuya defensa este propone el siguiente argumento:

Paso 1: Hay relaciones causales entre los fenómenos mentales y los fenómenos físicos.
Paso 2: Cada vez que hay sucesos con una relación de causa y efecto, esos sucesos deben estar sometidos a leyes causales estrictas y deterministas.
Paso 3: Pero no existen leyes causales estrictas y deterministas que relacionen lo mental y lo físico. En términos de Davidson, no hay leyes psicofísicas.

20 D. Davidson, "Mental Events", en L. Foster y J. W. Swanson (comps.), *Experience and Theory*, Amherst (Mass.), University of Massachusetts Press, 1970, reeditado en D. Davidson, *Essays on Actions and Events*, Oxford, Oxford University Press, 1980, pp. 207-227 [traducción española: "Sucesos mentales", en *Ensayos sobre acciones y sucesos*, Barcelona y México, Crítica/Instituto de Investigaciones Filosóficas de la UNAM, 1995].

Por lo tanto,
Paso 4: Conclusión. Todos los presuntos sucesos mentales son sucesos físicos.

Deben serlo para ejemplificar leyes físicas, y cuando los describimos como mentales, no hacemos sino elegir una categoría de sucesos físicos concordes con cierto vocabulario mental. Son sucesos mentales de acuerdo con una descripción, pero según otra también son físicos. El resultado, entonces, es una suerte de materialismo, un materialismo a cuyo entender el objeto de las ciencias psicológicas nunca podrá describirse mediante leyes universales como las vigentes en física, no porque se trate de un tipo misterioso de entidad espiritual o mental, sino porque las descripciones que utilizamos para caracterizarlo, las descripciones mentales, no se relacionan a la manera de una ley con los fenómenos físicos englobados en las descripciones físicas. El único argumento presentado por Davidson a favor de esta tesis es que los fenómenos mentales, por ejemplo las creencias y los deseos, están sujetos a restricciones de racionalidad, y esta "no tiene eco en física".

He intentado ser lo más justo posible en la exposición de las versiones clásicas del materialismo a lo largo del siglo pasado. Si no las hice ver atractivas en lo más mínimo, he fracasado en mi tarea de exponer las concepciones de otras personas. Debo confesar, sin embargo, que a mi entender todas esas teorías son desesperadamente inadecuadas. En ulteriores capítulos voy a analizar sus deficiencias. A los fines de la discusión inmediata, supondré que el conductismo no es una forma convincente de materialismo y que es preciso examinar las diferentes formas de fisicalismo, sobre todo el funcionalismo.

En el próximo capítulo, la mayor parte del análisis se referirá a la tradición histórica del funcionalismo que culmina en la inteligencia artificial fuerte. No diré nada sobre el monismo anómalo, porque puede incluirse bajo el encabezado general de las teorías de la identidad de casos. Ahora me ocuparé en forma breve, aunque espero que no injusta, del materialismo eliminativo. He mencionado tres argumentos en su apoyo. El primero dice que las entidades de la psicología popular se postulan como parte de una estructura teórica. Pero en general, eso no es cierto. Experimento directamente todos los procesos reales de pensamiento consciente relacionados con la decisión de intentar conseguir algo en función de mi deseo.

El segundo argumento es que, con toda probabilidad, las proposiciones de la psicología popular se demostrarán falsas. Sin embargo, si se estudia a los autores que sostienen esta opinión, se advertirá un inconveniente: son muy poco convincentes en su especificación de dichas proposiciones. A veces nos atribuyen creencias que sin lugar a dudas no sostenemos. Por ejemplo, un autor nos adjudica creer que si creemos *p* y que si *p* entonces *q*, creeremos *q*[21]. La afirmación es increíble. Implicaría, por ejemplo, que quien cree cada miembro de un conjunto complejo de proposiciones, *a*, *b*, *c*, etc., contenidas en las premisas de una prueba, donde las otras premisas aparecen en condicionales de la forma "si *a* entonces *d*", "si *b* entonces *e*", "si *c* entonces *f*", etc., cree automáticamente todas las consecuencias lógicas. ¡Si así fuera, esas pruebas lógicas y matemáticas tan complejas nunca podrían sorprendernos, porque habríamos creído la conclusión

21 P. M. Churchland, "Eliminative materialism...", *op. cit.*, p. 603.

desde el principio! El absurdo radica en confundir nuestro compromiso lógico con la verdad de una proposición con la creencia real en esta antes de conocer ese compromiso. Las pruebas lógicas y matemáticas complejas muestran lo que nuestra creencia en las premisas nos compromete a creer en la conclusión. No muestran que creíamos realmente esta última desde el comienzo.

Y, en rigor, los materialistas eliminativos vacilan en extremo a la hora de enunciar las proposiciones de la psicología popular. Creo que hay una razón para ello. Muchas de las proposiciones de la llamada psicología popular no son de hecho proposiciones empíricas. En cierto sentido son principios constitutivos, principios analíticos de nuestros contenidos mentales. Así, por ejemplo, aquí tenemos una proposición de la psicología popular: típicamente, las creencias pueden ser verdaderas o falsas. Ahora bien, el inconveniente de tratar esa proposición como si fuera una hipótesis susceptible de resultar falsa es que forma parte de la definición de la creencia: es un principio constitutivo. Es como decir que en el fútbol americano el *touchdown* vale seis puntos. La dificultad con que tropiezan los materialistas eliminativos radica en tratar las proposiciones de la denominada psicología popular como si fueran hipótesis empíricas, cosa que en muchos casos no son. Si leemos en el diario que investigadores del MIT han descubierto, mediante el uso de una tecnología informática de última generación, que el *touchdown* del fútbol americano no vale en realidad seis puntos sino 5,99999, sabemos que han cometido un error estúpido. La proposición de que el *touchdown* vale seis puntos forma parte de su definición misma, tal como aparece actualmente en las reglas del fútbol americano. No se puede

descubrir que es falsa del mismo modo que se descubre la falsedad en las proposiciones empíricas comunes. Algunos de los ejemplos de Churchland son así. El autor dice que, según una de las proposiciones de la psicología popular, quien teme *p* no quiere que *p* suceda. Pero si agregamos la cláusula "si todo lo demás sigue igual", la proposición forma parte de la definición del temor. Si temo algo y todo lo demás sigue igual, no quiero que la cosa temida suceda. En consecuencia, no se puede demostrar que las entidades psicológicas populares no existen mostrando en primer lugar que nuestras creencias sobre ellas son falsas, porque muchas de las proposiciones básicas de la psicología popular son de manera análoga principios definicionales, analíticos o constitutivos de las entidades de esa misma psicología. Por eso los esfuerzos de los enemigos de esta para refutarla son tan inapropiados. Esto no prueba que las entidades de la psicología popular existan, sino que un argumento planteado para demostrar su inexistencia no logra cobrar altura.

El último argumento contra la psicología popular es aún peor. La idea es que, como no podemos hacer una fluida reducción tipo-tipo de las creencias, los deseos, etc., a la neurobiología, de una manera u otra esas entidades, por lo tanto, no existen. Compárese, sin embargo, una proposición similar: no podemos hacer una fluida reducción tipo-tipo de los vehículos utilitarios deportivos, las raquetas de tenis o las casas de campo con pisos en desnivel a las entidades de la física atómica. Y no podemos hacerlo por razones implícitas en este capítulo: las raquetas de tenis, etc., son realizables de múltiples maneras en física. A decir verdad, la física atómica realmente no necesita la noción de vehículo utilitario deportivo, casa de campo con pisos en

desnivel o raqueta de tenis. Pero ¿cree alguien en su sano juicio que de ello se deduce que esas entidades no existen? Como argumento formal general, el hecho de que no logremos hacer reducciones tipo-tipo de una entidad a ciencias más básicas no demuestra que la entidad irreductible no exista. Todo lo contrario.

Hay una interesante ironía en todo este análisis. Los reduccionistas y los eliminativistas tienden a suponer que sus posiciones son muy diferentes. Los primeros creen que las entidades mentales existen pero se las puede reducir a sucesos físicos. Los segundos creen que dichas entidades no existen en absoluto. Pero una y otra posición equivalen prácticamente a la misma conclusión. Los reduccionistas dicen que no hay allí otra cosa que procesos cerebrales descritos de manera materialista. Los eliminativistas dicen que no hay allí otra cosa que procesos cerebrales descritos de manera materialista. La aparente diferencia es una diferencia de vocabulario. Los primeros materialistas querían mostrar que los estados mentales no existían como tales mostrando que podían sufrir una reducción tipo-tipo a las entidades de la neurobiología. Los ulteriores materialistas eliminativos querían mostrar que las entidades de la psicología del sentido común no existen en absoluto mostrando la imposibilidad de someterlas a una reducción tipo-tipo a las entidades de la neurobiología. Ninguno de los dos argumentos sirve, pero ambos sugieren que estas personas están resueltas a intentar demostrar que nuestras nociones corrientes de sentido común de lo mental no designan nada en el mundo real, y están dispuestas a proponer cualquier argumento que puedan imaginar en apoyo de esa conclusión.

CAPÍTULO 3

ARGUMENTOS CONTRA EL MATERIALISMO

En el capítulo anterior expuse parte de la historia del materialismo reciente y examiné los argumentos contra algunas de sus versiones, en especial el conductismo, la teoría de la identidad de tipos y el materialismo eliminativo. En este capítulo presentaré los argumentos más comunes contra el materialismo, concentrándome en el funcionalismo, porque es en la actualidad la versión más influyente de esa doctrina. En general, los ataques tienen la misma estructura lógica: la descripción materialista excluye algún rasgo esencial de la mente, como la conciencia o la intencionalidad. En la jerga de los filósofos, el análisis materialista omite proporcionar condiciones *suficientes* de los fenómenos mentales, porque es posible satisfacerlo sin contar con los fenómenos mentales apropiados. Estrictamente hablando, el funcionalismo no exige la adopción de una postura materialista. El funcionalista define los estados mentales en términos de relaciones causales y estas, en principio, podrían encontrarse en cualquier cosa. Tal como resultó el mundo, simplemente sucede que están en cerebros físicos, computadoras materiales y otros sistemas físicos. El análisis funcionalista presume de ser una verdad conceptual que analiza los conceptos mentales en términos causales. El hecho de que esas relaciones causales se realicen en el cerebro humano es un descubrimiento empírico, no una verdad conceptual. Pero la fuerza impulsora del funcionalismo fue un rechazo materialista del dualismo. Los funcionalistas quieren analizar los fenómenos mentales de una mane-

ra que evite toda referencia a algo intrínsecamente subjetivo y no físico.

I. Ocho argumentos (y medio) contra el materialismo

1. Qualia *ausentes*

Las experiencias conscientes tienen un aspecto cualitativo. En el hecho de tomar cerveza hay una sensación cualitativa muy distinta de la producida al escuchar la novena sinfonía de Beethoven. Varios filósofos estimaron útil introducir un término técnico para describir ese aspecto cualitativo de la conciencia. El término es *qualia*, cuyo singular en inglés es *quale**. Cada estado consciente es un *qualia*, porque en él existe cierta sensación cualitativa. Ahora bien, dice el antifuncionalista, el inconveniente del funcionalismo es que excluye los *qualia*. Desestima el aspecto cualitativo de nuestras experiencias conscientes, y por ello los *qualia* están ausentes de la descripción funcionalista. Los qualia tienen existencia real, de modo que cualquier teoría que la niegue, como lo hace el funcionalismo, es explícita o implícitamente falsa.

2. Inversión del espectro

Una serie de filósofos propusieron un planteamiento conexo, apoyado en un viejo experimento de

* En español suele utilizarse "qualia" de manera indistinta para el singular y el plural. De seguirse estrictamente el origen latino del término, el singular debería ser *qualis*. (N. del T.)

pensamiento vivido por mucha gente en la historia del tema, y también por muchas personas ajenas a la filosofía.

Supongamos que ni usted ni yo somos daltónicos. Ambos hacemos exactamente las mismas discriminaciones de colores. Si se nos pide que separemos los lápices rojos de los verdes, los dos elegiremos los rojos. Cuando el semáforo pasa de rojo a verde, ambos avanzamos sin demora. Pero supongamos también que, en realidad, nuestras experiencias internas son muy diferentes. Si yo pudiera tener la experiencia que usted llama "ver verde", la llamaría "ver rojo". Y de manera similar, si usted pudiera tener la experiencia que yo llamo "ver verde", la llamaría "ver rojo". Tenemos, en síntesis, una inversión entre rojo y verde. Esta pasa totalmente inadvertida para cualquier prueba comportamental, porque las pruebas identifican facultades de hacer discriminaciones entre objetos del mundo y no la capacidad de rotular experiencias internas. Estas últimas podrían ser diferentes aun cuando el comportamiento èxterno fuera exactamente el mismo. Pero si eso es posible, el funcionalismo no nos proporciona una descripción de la experiencia interna, porque esta queda al margen de toda explicación funcionalista. El funcionalista daría exactamente la misma descripción de mi experiencia y la suya, ambas caracterizadas por el enunciado "veo algo verde", pero como las experiencias son diferentes, el funcionalismo es falso.

3. Thomas Nagel: ¿cómo es ser un murciélago?

Uno de los primeros argumentos conocidos contra los tipos funcionalistas de materialismo fue pro-

puesto en un artículo de Thomas Nagel titulado "What Is It Like to Be a Bat?"[1] Según Nagel, el aspecto verdaderamente difícil del problema mente-cuerpo es la cuestión de la conciencia. Supongamos que tenemos una descripción funcionalista, materialista y neurobiológica plenamente satisfactoria de diversos estados mentales: creencias, deseos, esperanzas, temores, etc. De todos modos, esa descripción no explicará la conciencia. Nagel lo ilustra con el ejemplo de un murciélago. El estilo de vida de este animal es diferente del nuestro. Los murciélagos duermen todo el día colgados cabeza abajo de vigas y vuelan de un lado a otro durante la noche; se orientan mediante un sonar cuyas emisiones producen ecos al rebotar en objetos sólidos. Ahora bien, dice Nagel, alguien podría tener un conocimiento consumado de la neurofisiología del murciélago y de todos los mecanismos funcionales que le permiten vivir y orientarse; de todas maneras, algo quedaría excluido del conocimiento de esta persona: ¿cómo es ser un murciélago? ¿Cuál es la sensación de serlo? Y esa es la esencia de la conciencia. En todo ser consciente hay un aspecto "cómo es ser" de su existencia. Y ese aspecto queda al margen de cualquier descripción objetiva de la conciencia porque una descripción objetiva no puede explicar el carácter subjetivo de esta.

1 T. Nagel, "What Is It Like to Be a Bat?", *Philosophical Review*, 83, 1974, pp. 435-450, reeditado en D. Chalmers (comp.), *The Philosophy of Mind: Classical and Contemporary Readings*, Nueva York, Oxford University Press, 2002 [traducción española: "¿Cómo es ser un murciélago?", en Obeth Hansberg y Maite Ezcurdia (comps.), *La naturaleza de la experiencia*, 1, *Sensaciones*, México, Instituto de Investigaciones Filosóficas de la UNAM, 2003].

4. Frank Jackson: lo que Mary no sabía

El filósofo australiano Frank Jackson propuso un argumento similar[2]. Este autor imagina a una neurobióloga, Mary, que sabe todo lo que puede saberse sobre la percepción del color. Tiene un conocimiento completo y acabado de la neurofisiología de nuestro aparato perceptivo del color, así como de la física de la luz y el espectro de los colores. Pero, dice Jackson, imaginemos que fue criada en un ambiente absolutamente blanco y negro. Mary nunca ha visto nada de color: sólo blanco, negro y matices de gris. Ahora bien, dice nuestro autor, parece evidente que algo ha quedado fuera de su conocimiento. Lo que queda afuera es, por ejemplo, la apariencia real del color rojo. Pero de ser así, una descripción funcionalista o materialista de la mente excluiría algo, porque una persona puede tener un conocimiento completo de todo lo que es posible saber de acuerdo con una descripción de esas características, sin saber cómo son los colores. Y el problema de los colores es sólo un caso especial del problema de las experiencias cualitativas en general. Toda descripción de la mente que deje al margen esas experiencias cualitativas es inadecuada.

2 F. Jackson, "What Mary Didn't Know", *Journal of Philosophy*, 83, 1982, pp. 291-295, reeditado en T. O'Connor y D. Robb (comps.), *Philosophy of Mind*, Nueva York, Routledge, 2003 [traducción española: "Lo que María no sabía", en O. Hansberg y M. Ezcurdia (comps.), *La naturaleza de la experiencia*, 1, *op. cit.*], y "Epiphenomenal Qualia", *Philosophical Quarterly*, 32, 1986, pp. 127-136, reeditado en D. Chalmers, *The Philosophy of Mind...*, *op. cit.* [traducción española: "Qualia epifenoménicos", en O. Hansberg y M. Ezcurdia (comps.), *La naturaleza de la experiencia*, 1, *op. cit.*]

5. *Ned Block: la nación china*

Ned Block propuso un quinto argumento en respaldo de la misma concepción general antifuncionalista[3]. Block dice que podríamos imaginar una gran población que cumpliera los pasos de un programa funcionalista del tipo presuntamente ejecutado por el cerebro. Así, por ejemplo, imaginemos que hay mil millones de neuronas en el cerebro y mil millones de habitantes en China. (La cifra de mil millones de neuronas es, desde luego, absurdamente pequeña para el cerebro, pero no tiene importancia para este argumento.) Ahora podríamos suponer que así como el cerebro cumple ciertos pasos funcionalistas, podemos lograr que la población de China haga exactamente lo mismo. No por ello, sin embargo, los chinos tendrán como conjunto algún estado mental, a diferencia del cerebro que *sí* los tiene.

6. *Saul Kripke: designadores rígidos*

Saul Kripke presentó un argumento puramente lógico contra todas las versiones de la teoría de la identidad[4]. Ese argumento apela al concepto de "designador rígido", definido como una expresión que siempre se refiere al mismo objeto en cualquier situación posible. Así, la expresión "Benjamin Franklin" es un designador rígido porque en el uso que ahora invoco siempre se refiere a la misma persona. Esto no significa decir, por supuesto, que yo no pueda bautizar a mi perro como

3 N. Block, "Troubles with Functionalism", *op. cit.*

4 S. A. Kripke, *Naming and Necessity*, Cambridge (Mass.), Harvard University Press, 1980, partes importantes del cual se reeditaron en D. Chalmers, *The Philosophy of Mind...*, *op. cit.*, pp. 329-332 [traducción española: *El nombrar y la necesidad*, México, UNAM, 1996].

"Benjamin Franklin", pero en este caso hay un uso y un significado diferentes de la expresión. Según el significado convencional, "Benjamin Franklin" es un designador rígido. En cambio, la expresión "el inventor de la hora de verano", aunque también se refiere a Benjamin Franklin, no es un designador rígido, porque resulta fácil imaginar un mundo en el cual Franklin no fuera el inventor del aprovechamiento de la luz solar en verano. Tiene sentido decir que otra persona y no el verdadero inventor podría haber inventado la hora de verano, pero no lo tiene decir que otro individuo al margen de Benjamin Franklin podría haber sido Benjamin Franklin. Por esas razones, "Benjamin Franklin" es un designador rígido, mientras que "el inventor de la hora de verano" es un designador no rígido.

Previsto de la noción de designadores rígidos, Kripke procede a examinar las proposiciones de identidad. Su tesis es que estas, en las cuales un término es rígido y el otro no lo es, no son en general necesariamente verdaderas; podrían resultar falsas. Así, la frase "Benjamin Franklin es idéntico al inventor de la hora de verano" es verdadera, pero sólo de manera contingente. Podemos imaginar un mundo en el cual sea falsa. Sin embargo, dice Kripke, cuando ambas partes de la proposición son rígidas, esta última, si es verdadera, debe serlo por necesidad. Por lo tanto, la proposición "Samuel Clements es idéntico a Mark Twain" es necesariamente verdadera, porque no puede haber un mundo en el cual existan uno y otro pero sean dos personas diferentes. Ocurre otro tanto con las palabras que nombran tipos de cosas. El agua es lo mismo que H_2O, y como ambas identidades son rígidas, la identidad debe ser necesaria. Y aquí está la relevancia para el problema mente-cuerpo: si en el lado izquierdo de nuestra pro-

posición de identidad tenemos una expresión referida en forma rígida a un tipo de estado mental, y en el lado derecho una expresión también rígidamente referida a un tipo de estado cerebral, la proposición, si fuera verdadera, debería serlo por necesidad. Entonces, si el dolor fuera realmente idéntico a las estimulaciones de las fibras C, la proposición "dolor = estimulaciones de las fibras C", de ser verdadera, tendría que serlo de manera necesaria. Pero resulta evidente que no es necesariamente verdadera. En efecto, aunque hubiera una correlación estricta entre dolores y estimulaciones de las fibras C, de todos modos sería fácil imaginar la existencia de un dolor sin que hubiera ninguna estimlación de ese tipo, así como la existencia de una estimulación de las fibras C sin su correspondiente dolor. Pero en ese caso la proposición de identidad no es necesariamente verdadera, y si no lo es no puede ser verdadera en absoluto. Por lo tanto, es falsa. Y lo que vale para la identificación de los dolores con sucesos neurobiológicos vale para cualquier identificación entre estados mentales conscientes y sucesos físicos.

7. *John Searle: la habitación china*

Un argumento explícitamente dirigido contra la inteligencia artificial fuerte fue presentado por el autor de este libro[5]. La estrategia del planteamiento consiste

5 J. R. Searle, "Minds, Brains and Programs", *Behavioral and Brain Sciences*, 3, 1980, pp. 417-424, reeditado en muchas publicaciones, entre ellas T. O'Connor y D. Robb (comps.), *Philosophy of Mind*, *op. cit.*, pp. 332-352 [traducción española: "Mentes, cerebros y programas", en Margaret A. Boden (comp.), *Filosofía de la inteligencia artificial*, México, Fondo de Cultura Económica, 1994].

en apelar a nuestras experiencias de primera persona para someter a prueba cualquier teoría de la mente. Si la inteligencia artificial fuerte fuera verdadera, todo el mundo debería poder adquirir cualquier capacidad cognitiva a través de la mera ejecución del programa informático que la simula. Probemos esta idea con el chino. De hecho, no entiendo en absoluto ninguna de sus variantes. Ni siquiera puedo diferenciar la escritura china de la escritura japonesa. Imaginemos, empero, que estoy encerrado en una habitación con cajas llenas de símbolos chinos y tengo un libro de instrucciones; en sustancia, un programa informático que me permite responder preguntas dirigidas a mí en chino. Me presentan símbolos que, desconocidos para mí, son preguntas; consulto el libro para saber qué debo hacer; selecciono símbolos de las cajas, los manipulo de acuerdo con las instrucciones del programa y dispongo los requeridos, que se interpretan como respuestas. Podemos suponer que paso la prueba de Turing sobre la comprensión del chino, pero, de todas maneras, no entiendo una palabra de ese idioma. Y si no entiendo chino a partir de la ejecución del programa informático apropiado, tampoco lo hará ninguna otra computadora sobre la mera base de implementar el programa, porque ninguna computadora tiene algo que yo no tenga.

Podrá verse la diferencia entre computación y comprensión real si se imagina cómo es también para mí contestar preguntas en inglés. Supongamos que en el mismo cuarto me dan preguntas en esa lengua y debo responderlas. Desde afuera, mis respuestas a las preguntas en inglés y en chino son igualmente buenas. Paso la prueba de Turing para ambos idiomas. Pero desde adentro hay una enorme diferencia. ¿Cuál es, exactamente? En inglés entiendo el significado de las

palabras; en chino no entiendo nada. En chino sólo soy una computadora.

El argumento de la habitación china fue un golpe directo al corazón del proyecto de la inteligencia artificial fuerte. Con anterioridad a su publicación, los ataques a la inteligencia artificial solían decir que la mente humana dispone de ciertas habilidades que la computadora no tiene y jamás podría tener[6]. Esta estrategia siempre es peligrosa, porque tan pronto como alguien dice que hay determinada clase de tareas que las computadoras son incapaces de hacer, surge la muy fuerte tentación de diseñar un programa que lleve a cabo precisamente eso. Y esto ha sucedido con frecuencia. Cuando ocurre, los críticos de la inteligencia artificial suelen decir que la tarea, de todos modos, no era tan importante y los éxitos informáticos en verdad no cuentan. Los defensores de la inteligencia artificial sienten, con alguna justificación, que les mueven constantemente la línea de llegada. El argumento de la habitación china adoptó una estrategia enteramente diferente. Supone un completo éxito de la inteligencia artificial en la simulación de la cognición humana. Supone que los investigadores de la disciplina pueden diseñar un programa que pase la prueba de Turing para la comprensión del chino o cualquier otra cosa. De todos modos, en lo concerniente a la cognición humana, esos logros son sencillamente irrelevantes. Y lo son por una razón profunda: la computadora opera a través de la manipulación de símbolos. Sus procesos se definen de manera puramente sintáctica, mientras que la mente

6 H. Dreyfus, *What Computers Can't Do*, edición revisada, Nueva York, Harper & Row, 1979.

humana tiene algo más que símbolos no interpretados: asocia significados a ellos.

Hay un desarrollo adicional del planteamiento que me parece más convincente, aunque se le prestó mucha menor atención que al argumento original de la habitación china. En este último, yo suponía que la atribución de sintaxis y capacidad de cómputo al sistema no era problemática. Pero si lo pensamos mejor veremos que *el cómputo y la sintaxis son relativos al observador*. Salvo en los casos en que una persona hace cálculos concretos en la mente, en la naturaleza no hay cómputos intrínsecos ni originales. Cuando sumo dos más dos para llegar a cuatro, ese cálculo no es relativo al observador. Lo hago con prescindencia de lo que cualquiera piense. Pero cuando tecleo "2 + 2" en mi calculadora de bolsillo y esta muestra un "4", la máquina no sabe nada de computación, aritmética o símbolos, porque no sabe nada de nada. Intrínsecamente se trata de un circuito electrónico complejo que *usamos* para calcular. Las transiciones de estados eléctricos son inherentes a la máquina, pero el cálculo está en los ojos del espectador. Lo que vale para la calculadora vale para cualquier computadora comercial. El cómputo está en la máquina como la información está en el libro. Está en ella, de acuerdo, pero es relativo al observador y no intrínseco. Por esa razón, no podríamos descubrir que el cerebro es una computadora digital, porque la computación no se descubre en la naturaleza, se le asigna a esta. De modo que la pregunta: ¿es el cerebro una computadora digital?, está mal formulada. Si trata de averiguar si el cerebro es intrínsecamente una computadora digital, la respuesta es que nada lo es intrínsecamente, excepto para agentes conscientes que piensan por medio de cómputos. Si la pregunta es: ¿podríamos

asignar una interpretación computacional al cerebro?, la respuesta es que podemos atribuirla a cualquier cosa.

No desarrollo el argumento aquí, pero quiero que el lector conozca al menos su esqueleto. En el capítulo 9 de *The Rediscovery of the Mind* se encontrará una exposición más completa[7].

8. *La concebibilidad de los zombis*

Uno de los más antiguos argumentos, y en cierto modo el argumento subyacente a varios otros, es el siguiente: es concebible que pueda haber un ser que en el aspecto físico sea exactamente igual a mí en todo, pero carezca por completo de vida mental. Una de las versiones del argumento plantea la posibilidad lógica de que exista un zombi igual a mí molécula por molécula, pero sin vida mental alguna. En filosofía, un zombi es un sistema que se comporta como los seres humanos pero no tiene vida mental, conciencia o intencionalidad real; y este argumento afirma que los zombis son posibles desde el punto de vista de la lógica. Y si lo son, es decir, si es lógicamente posible que un sistema tenga el comportamiento y los mecanismos funcionales apropiados, e incluso la estructura física pertinente, y pese a ello carezca de vida mental, los análisis conductistas y funcionalistas están equivocados. No enuncian las condiciones lógicamente suficientes para tener una mente.

Este argumento se presenta en varias formas. Uno de sus primeros enunciados contemporáneos es el de Thomas Nagel[8]. Este autor sostiene:

7 J. R. Searle, *The Rediscovery of the Mind*, *op. cit.*

8 T. Nagel, "Armstrong on the Mind", en N. Block (comp.), *Readings in Philosophy of Psychology*, *op. cit.*, p. 205.

> Puedo concebir que mi cuerpo haga precisamente lo que hace ahora, adentro y afuera, con una total causación física de su comportamiento (incluyendo el comportamiento típicamente autoconsciente), pero sin ninguno de los estados mentales que experimento en estos momentos y, para el caso, sin ningún otro. Si esto es realmente concebible, los estados mentales deben ser distintos del estado físico del cuerpo.

Este planteamiento es una especie de imagen especular del argumento de Descartes. El filósofo sostenía la posibilidad de concebir la existencia de mi mente sin mi cuerpo, por lo cual la primera no podía ser idéntica al segundo. Y el argumento de Nagel dice que es concebible que mi cuerpo exista y sea exactamente tal como es, pero sin la mente; por lo tanto, esta no es idéntica a él ni a ninguna de sus partes u operaciones.

9. *La forma aspectual de la intencionalidad*

Sólo puedo presentar el argumento final en una forma abreviada (por lo cual lo califico de medio argumento), porque aún no he explicado la intencionalidad en detalle suficiente como para exponerlo en su totalidad. Me parece, no obstante, que puedo dar una idea bastante clara de su planteamiento. Los estados intencionales, como las creencias y los deseos, representan el mundo según algunos aspectos y dejando de lado otros. Por ejemplo, el deseo de agua no es igual al deseo de H_2O, porque una persona podría querer agua sin saber que es H_2O e incluso creyendo que no lo es. Como todos los estados intencionales representan según distintos aspectos, podríamos decir que tienen una forma aspectual. Pero una descripción causal de la intencio-

nalidad como la proporcionada por los funcionalistas no puede captar diferencias en la forma aspectual porque la causalidad carece de esta. Todo lo que el agua causa, el H_2O lo causa; y la causa del agua, cualquiera sea, es la causa del H_2O. El análisis funcionalista de mi creencia de que esta sustancia es agua y de mi deseo de agua presentados en términos causales no puede distinguir una y otro de mi creencia de que la sustancia es H_2O y mi deseo de H_2O. Pero son claramente distintos, y por lo tanto el funcionalismo fracasa.

Y no es posible responder a este argumento diciendo que podemos preguntar a la persona: "¿Cree usted que esta sustancia es agua? ¿Cree que esta sustancia es H_2O?", porque el problema que teníamos con respecto a la creencia y el deseo surge ahora en relación con el *significado*. ¿Cómo sabemos que la persona en cuestión se refiere con "H_2O" y con "agua" a lo mismo que nosotros llamamos "H_2O" y "agua"? Si nuestras únicas guías son el comportamiento y las relaciones causales, no tenemos elementos suficientes para distinguir distintos significados en la cabeza del agente. En suma, las traducciones alternativas e incongruentes serán congruentes con todos los datos causales y comportamentales[9].

9 La insuficiencia del comportamiento para discriminar entre significados discriminables fue demostrada por W. V. O. Quine, *Word and Object*, Cambridge (Mass.), Harvard University Press, 1962 [traducción española: *Palabra y objeto*, Barcelona, Herder, 2001]. Quine no consideraba que el argumento fuera una *reductio ad absurdum* de las descripciones conductistas del significado. En J. R. Searle, "Indeterminacy, Empiricism, and the First Person", *Journal of Philosophy*, 84(3), marzo de 1987, pp. 123-147, reeditado en J. R. Searle, *Consciousness and Language*, Cambridge, Cambridge University Press, 2002, se encontrará una crítica de las concepciones de Quine.

No he visto ninguna formulación anterior de este argumento, que sólo se me ocurrió al escribir el presente libro. Para resumirlo en la jerga que explicaré en el capítulo 6, la intencionalidad implica en esencia una forma aspectual. Toda representación mental se muestra en aspectos representacionales. La causación también tiene aspectos, pero no son representacionales. Los conceptos mentales no se pueden analizar en términos causales porque la forma aspectual representacional de lo intencional se pierde en la traducción. Por eso los enunciados acerca de la intencionalidad son intensionales con s, pero los enunciados sobre la causación, de la forma A causó B, son extensionales. (No se preocupe si no entiende este párrafo. Ya llegaremos a esto en el capítulo 6).

II. Respuestas materialistas a los argumentos precedentes

No es de sorprender que los defensores del funcionalismo, la teoría de la identidad y la inteligencia artificial fuerte se sientan en general capaces de dar una respuesta a los argumentos antes mencionados (con excepción del último, que se publica aquí por primera vez). Hay una abundante literatura sobre el tema, y no intentaré revisarla en este libro. (Conozco más de cien ataques publicados contra el argumento de la habitación china sólo en inglés, y supongo que debe haber decenas más que ignoro, tanto en ese idioma como en otros.) Pero algunos de los argumentos en defensa del materialismo son muy comunes y han conquistado vasta aceptación, por lo cual vale la pena discutirlos aquí.

Respuestas a Nagel y Jackson

Una respuesta clásica dada por los materialistas contra Nagel y Jackson fue la siguiente: ambos argumentos se apoyan en lo conocido, sea lo que alguien podría conocer sobre la fisiología del murciélago o lo que Mary conoce acerca de la fisiología de la percepción. Así, uno y otro afirman que aun un conocimiento perfecto de los fenómenos funcionales o fisiológicos de tercera persona dejaría algo al margen. Excluiría los fenómenos experienciales subjetivos y cualitativos de primera persona. La respuesta a esa objeción es que cualquier argumento basado en lo que se conoce de acuerdo con una descripción y no se conoce de acuerdo con otra, es insuficiente para establecer la inexistencia de una identidad entre las cosas descritas por ambas. De tal modo, para considerar un ejemplo obvio, supongamos que Sam sabe que el agua es húmeda, pero no sabe que el H_2O lo es. Supongamos, para continuar, que alguien sostiene que el agua no puede ser idéntica al H_2O porque en este último hay algo que Sam no sabe, pero sí sabe acerca de la primera. A mi juicio, todos pueden darse cuenta de que el argumento es malo. El hecho de que uno pueda saber algo sobre una sustancia de acuerdo con una descripción, por ejemplo como agua, y no saber eso mismo acerca de ella según otra descripción, por ejemplo como H_2O, no implica que el agua no sea H_2O.

¿Será válido este argumento contra Nagel y Jackson? Para plantear un caso paralelo, habría que argumentar del siguiente modo. Mary sabe, por ejemplo, que el proceso neuronal x437B es causado por los objetos rojos, pero ignora que este tipo de experiencia del rojo es causada por esos mismos objetos. Lo ignora porque nunca tuvo la experiencia del color rojo. Y la supuesta conclusión es que esa experiencia del color no puede ser idéntica a los procesos x437B. Este argumento es tan falaz como el que consideramos antes con referencia al agua y el H_2O. Y si Nagel y Jackson pretendieran que sus argumentos se interpretaran de esa manera, podría acusárselos de ser igualmente falaces.

¿Refutamos así el punto de vista de ambos autores? Creo que no. Es posible enunciarlo como un argumento sobre el conocimiento, y Nagel y Jackson suelen plantearlo de esa forma (en efecto, la tesis del segundo se denomina con frecuencia "argumento del conocimiento"), pero en su significado no está sujeto a la acusación de cometer la falacia de suponer que, si se conoce algo de una entidad de acuerdo con una descripción y se desconoce sobre otra entidad según otra descripción, la primera entidad no puede ser idéntica a la segunda. El quid del argumento no radica en apelar a la ignorancia del especialista en murciélagos o de Mary. Consiste en decir que existen fenómenos reales que quedan necesariamente al margen del alcance de su conocimiento, mientras este sólo se refiera a hechos físicos objetivos y de tercera persona. Los fenómenos reales son las sensaciones de los murciélagos y las experiencias del color, respectivamente, y se trata de fenómenos subjetivos conscientes y de primera persona. En el caso de Mary, el problema no es sólo que carece de *información* acerca de algunos otros fenómenos, sino que hay un tipo determinado de *experiencia* que ella aún no ha vivido. Y esa experiencia, un fenómeno subjetivo de primera persona, no puede ser idéntica a los correlatos neuronales y funcionales objetivos y de tercera persona. La cuestión epistemológica, la información, no es sino una manera de llegar a la diferencia ontológica subyacente. Observaciones similares son

válidas para el ejemplo del murciélago de Nagel. El problema no es que el investigador carezca de información; a decir verdad, puede tener una información perfecta de tercera persona. Lo que le falta es la experiencia vivida por el murciélago, el tipo de fenómeno producido en la conciencia de este. Por lo tanto, aunque ambos argumentos se enuncian como si fueran epistémicos, en realidad creo que, apropiadamente interpretados, son ontológicos y, entonces, no están sujetos a la objeción que considerábamos.

La forma lógica de los argumentos es esta: me sitúo en una relación con ciertas entidades, mis experiencias de los colores. Y el murciélago se sitúa en una relación con ciertas entidades, sus experiencias de lo que es ser un murciélago. Una descripción completa de tercera persona del mundo deja al margen esas entidades y por lo tanto es incompleta. Los ejemplos de Mary y el especialista en murciélagos son un modo de ilustrar la incompletitud.

El verdadero problema de todas las formas de reduccionismo, como veremos, es que se enfrentan a la siguiente cuestión: ¿hay dos fenómenos o sólo uno? En el caso del agua hay realmente un solo fenómeno. El agua consiste en su totalidad de moléculas de H_2O; hay una sola cosa, agua, exclusivamente compuesta de esas moléculas. Pero cuando se trata de identificar rasgos de la mente, como la conciencia y la intencionalidad, con rasgos del cerebro, como los estados computacionales o los estados neurobiológicos, parece que debería haber dos características, porque los fenómenos mentales tienen una ontología de primera persona, en cuanto sólo existen si son experimentados por algún sujeto humano o animal, algún "yo" ["*I*"] que vive la experiencia. Y esto los hace irreducibles a toda ontología de tercera

persona, todo modo de existencia que sea independiente de un agente experienciador. El hincapié en la diferencia entre la ontología de primera persona y la de tercera persona es el verdadero sentido de todos esos argumentos contra este tipo de reduccionismo.

Respuestas a Kripke con respecto a los designadores rígidos

Una respuesta habitual al argumento de Kripke concerniente a los designadores rígidos es que no refuta las proposiciones de la identidad de casos[10]. La idea es que el argumento podría ser válido contra las identidades de tipos, pero no contra las identidades de casos. Así pues, aun cuando sea posible imaginar, en general, la activación de una fibra C sin un dolor y un dolor sin la activación de una fibra C, en esta instancia específica, en este caso particular, yo no podría sobrellevar la activación de esa misma fibra sin sentir dolor ni sentir ese mismo dolor sin sufrir dicha activación. ¿Responde esto al argumento de Kripke? No veo de qué modo. Si se me concede que la experiencia tiene efectivamente dos rasgos, la sensación de dolor y la activación de las fibras C, el argumento de Kripke parece ser valedero. Yo podría haber tenido la misma sensación sin que hubiese ninguna activación correlacionada de fibras C, y esa misma activación sin el correlato de ninguna sensación. Ahora bien, siempre es posible, desde luego, emparchar las cosas y limitarse a establecer un criterio para la identidad de la sensación y la activación de las fibras, la coocurrencia de ambas. Así, si el dolor

10. C. McGinn, "Anomalous Monism and Kripke's Cartesian Intuitions", en N. Block (comp.), *Readings in Philosophy of Psychology*, *op. cit.*, pp. 156-158.

es en parte lo que es porque coocurre con la activación de las fibras C, y esta última es en parte lo que es porque coocurre con ese dolor, llegamos a una necesaria identidad entre uno y otra. Sin embargo, todavía no hemos alcanzado la meta de la identidad de casos, porque ahora tenemos una versión del dualismo de las propiedades. Lo que decimos es que una y la misma entidad tiene a la vez propiedades objetivas de activación de las fibras C y propiedades subjetivas dolorosas. Volveré a este punto en el capítulo 4.

En realidad, no está verdaderamente muy claro en qué medida utilizamos correlaciones, e incluso correlaciones causales, como condiciones de identidad de las sensaciones. Supongamos que siento un dolor; supongamos, además, que ese dolor tiene una causa específica. Imaginemos, sin embargo, que mientras siento ese mismo dolor, la experiencia continúa pero la causa inicial desaparece para dejar su lugar a otra. ¿Diremos que he tenido dos dolores diferentes porque, si bien había una sensación continua, las causas eran dos? ¿O diremos que tuve un solo dolor continuo, pero con una causa en su primera parte y otra en la segunda? No creo que el lenguaje común y corriente nos resuelva esta cuestión. Debemos tomar una decisión. Lo importante, empero, es ver que, en el caso de los dolores, es preciso distinguir entre la experiencia real por un lado y el sustrato neurobiológico por otro. No puedo decirles cuánta resistencia oponen los filósofos materialistas a esta observación evidente.

Respuestas al argumento de la habitación china de Searle

Retomo de mala gana la discusión sobre la habitación china porque ya la he discutido en muchos luga-

res. No obstante, a los efectos de este libro, vale la pena señalar las deficiencias de las objeciones habituales planteadas contra dicho argumento. Para mi sorpresa, el argumento convencional contra la habitación china es lo que denomino "réplica de los sistemas". La idea de esta réplica es que aunque no entienda chino, el hombre encerrado en la habitación es sólo una parte de un sistema más amplio formado por el cuarto, los libros de instrucciones, las ventanas, las cajas, el programa, etc. El que entiende esa lengua no es el hombre, sino todo el sistema. Como me dijo alguien, la habitación entera entiende chino. Es importante decir exactamente por qué esta réplica es inadecuada. Si preguntamos: ¿por qué no entiendo chino en la habitación?, la respuesta es: porque no tengo manera de conocer el significado de ninguno de los símbolos chinos. Tengo la sintaxis pero no la semántica. Pero entonces, si no tengo posibilidad de pasar de la sintaxis a la semántica, tampoco la habitación entera la tiene. No posee los recursos de que yo carezco para asociar significados a símbolos. Ilustré la situación con una ampliación del experimento de pensamiento. Imaginemos que me libero de la habitación y trabajo al aire libre. Hago todos los cálculos mentalmente y memorizo el programa y la base de datos. Podemos incluso imaginar que trabajo en campo abierto. De todas maneras, sigue siendo imposible que entienda chino y tampoco hay en mí subsistemas o rasgos capaces de entenderlo, porque no hay nada en mí, ni en ninguno de mis subsistemas, ni en ninguno de los sistemas más amplios de los cuales formo parte, que permita al sistema asociar significados a los símbolos. Manipular los símbolos es una cosa, conocer sus significados es otra. Las computadoras se definen en términos de manipulación simbólica, y esta,

por sí misma, no es ni constitutiva del significado ni suficiente para conocerlo.

La distinción entre sintaxis y semántica es tan importante para el resto del argumento de este libro que quiero decir algo más sobre ella aquí. Para que pueda haber comunicación lingüística humana, debe haber un lenguaje. Un lenguaje consiste de símbolos, por lo común palabras, combinadas en oraciones. Estos elementos: símbolos, palabras, oraciones, son sintácticos. Pero el lenguaje sólo funciona si son significativos: si tienen significado. ¿Qué es, empero, el significado? La literatura filosófica, lingüística y psicológica da muchas definiciones diferentes. Tengo opiniones bien claras sobre cuáles son acertadas y cuáles incorrectas, pero a los efectos de este argumento esas discrepancias no importan. Cualquier definición juiciosa del significado debe reconocer la distinción entre los símbolos, conceptualizados como entidades sintácticas puramente abstractas, y los significados asociados a ellos. Los símbolos deben distinguirse de sus significados. Por ejemplo, si escribo una frase en alemán: *"Es regnet"*, el lector verá palabras en la página y, por lo tanto, verá los objetos sintácticos, pero si no sabe alemán sólo advertirá la sintaxis y no la semántica. Se encontrará en la situación en que me encuentro cuando estoy en la habitación china, donde conozco la sintaxis del sistema computacional, pero no sé qué significa nada de eso.

Respuestas a la concebibilidad de los zombis

Hay muchos análisis del argumento de los zombis. Una respuesta consiste en negar lisa y llanamente la posibilidad de concebir zombis que se comporten como nosotros pero carezcan de vida mental. La estrategia no parece muy prometedora, porque desde un

punto de vista intuitivo es muy fácil, en apariencia, imaginar una máquina que sea exactamente como yo, pero sin conciencia. Daniel Dennett[11] apoya la estrategia con la siguiente analogía. Supongamos que alguien ha señalado la existencia de barras de hierro que en todos los aspectos se comportan exactamente igual que los imanes [*magnets*], pero no son imanes sino zagnetos [*zagnets*]. Eso es inconcebible porque, dice Dennett, los zagnetos serían simplemente imanes. De manera análoga, una máquina que se comporta en todos los sentidos como un agente consciente es un agente consciente. Los zagnetos son imanes y los zombis son agentes conscientes.

Esta analogía no funciona. Una descripción apropiada de un zagneto *implicará* que es un imán, pero ninguna descripción de tercera persona de un sistema físico implicará que este tiene estados conscientes porque hay dos fenómenos diferentes, las estructuras neurobiológicas comportamentales y funcionales de tercera persona y la experiencia consciente de primera persona.

A veces se da otra respuesta al argumento de los zombis: si fuera correcto, la conciencia se convertiría en un epifenómeno. Si pudiéramos exhibir el mismo comportamiento sin conciencia, significaría que esta no hace trabajo alguno. Esa respuesta descansa sobre un malentendido. El quid del argumento de los zombis es mostrar que la conciencia, por un lado, y el comportamiento y las relaciones causales, por otro, son fenómenos diferentes, para lo cual se demuestra la posibilidad lógica de tener uno sin otro. Pero esa posibilidad lógi-

11 D. Dennett, "Back from the Drawing Board", en D. Dahlbom, *Dennett and His Critics*, Cambridge (Mass.), Blackwell, 1993, p. 211.

ca no implica que la conciencia no cumpla ningún papel en el mundo real. De manera análoga: la combustión de la gasolina no es lo mismo que el movimiento del automóvil, porque es concebible tener una cosa sin la otra. Pero la existencia de la posibilidad lógica de que los autos se muevan sin gasolina, e incluso sin combustible alguno, no muestra que la gasolina y otros combustibles sean epifenómenos.

III. Conclusión

¿Qué deberíamos decir de estos argumentos? En filosofía siempre es importante dar un paso atrás y observar las cuestiones desde una perspectiva intelectual e histórica más amplia. ¿Por qué tantos filósofos se sienten en la obligación de negar ciertas afirmaciones de sentido común, por ejemplo que tenemos efectivamente pensamientos y sentimientos conscientes; que tenemos verdaderos estados intencionales tales como creencias, esperanzas, temores y deseos; que esos estados intencionales son causados por procesos localizados en el cerebro y funcionan a su vez de manera causal, y que son partes intrínsecas reales del mundo real y participan de nuestra vida biológica del mismo modo que la digestión, el crecimiento o la secreción de bilis? La respuesta debe buscarse en la historia. En conjunto, los fracasos del dualismo y el éxito de las ciencias físicas nos inducen a pensar que, de una u otra manera, debemos ser capaces de presentar una descripción de todo lo susceptible de decirse del mundo real en términos completamente materialistas. La existencia de algunos fenómenos *mentales* irreducibles no encaja y parece repulsiva en el plano intelectual. Es indigerible. Adviértase que la gente no tiene estos problemas en lo concerniente a otras partes de nuestra vida biológica. Nadie siente la necesidad de reducir otros fenómenos biológicos a alguna otra cosa. Nadie cree, por ejemplo, que la existencia de los pulgares plantee algún problema y sea preciso someterlos a un análisis funcionalista para mostrar que pueden definirse por entero desde el punto de vista de nuestra conducta prensil. Si los filósofos se preocupan por los dolores y no por los pulgares, es porque los primeros, según la visión del sentido común, tienen una especie de componente cualitativo irreductiblemente privado y subjetivo, y la meta es siempre liberarse de ese tipo de cosas.

En la historia que hemos examinado se hizo una distinción entre conciencia e intencionalidad. Muchos filósofos habrían estado dispuestos a coincidir en que no había ninguna descripción funcionalista de la conciencia, pero pretendían sostener que la intencionalidad estaba sometida a una reducción funcionalista y que la descripción computacional de la mente nos mostraba una reducción hermosa y científicamente impecable. Olvídense de la conciencia, que de todas maneras no tiene relevancia científica. Lo importante de la mente es su capacidad para el procesamiento de información, y la computadora moderna nos da por fin el modelo adecuado para comprender las capacidades de la mente en esa materia. Esta concepción del materialismo moderno, según la cual la conciencia puede hacerse a un lado mientras nos concentramos en la intencionalidad, explica por qué la habitación china sufrió tantos ataques más que otros argumentos. Ese argumento, en efecto, amenazaba la ciudadela misma de la descripción funcionalista computacional, que es la idea de que si tuviéramos las relaciones adecuadas de entrada y salida y contáramos con el programa pertinente que media

entre ellas, tendríamos todo el contenido de la intencionalidad. El argumento de la habitación china muestra que en el ser humano suceden dos cosas: la primera, los símbolos concretos de los cuales el hombre es consciente cuado piensa; la segunda, el significado, interpretación o sentido que se asocia a ellos.

Ahora bien, siempre está el problema de la reducción. ¿Hay dos fenómenos o sólo uno? Si los fenómenos reales son dos, no hay modo de negar la existencia de uno sin incurrir en una falsedad; no hay manera de hacer una reducción ontológica de uno a otro.

Entonces, ¿en qué situación quedamos? ¿Estamos obligados a volver al dualismo? Si el materialismo no ha logrado enunciar una alternativa convincente al dualismo tradicional en cuyo reemplazo se postulaba, ¿por qué no regresar a este último? Y, a decir verdad, ¿no admitimos tácitamente el dualismo cuando decimos que la conciencia y la intencionalidad son irreducibles?

Creo que nuestros verdaderos problemas tienen que ver con una maraña de confusiones conceptuales que trataré de dilucidar en el próximo capítulo.

Terminamos este capítulo en un estado intelectual depresivo: ni el dualismo ni el materialismo son aceptables, no obstante lo cual se nos presentan como las únicas posibilidades. Por otra parte, sabemos de manera independiente que lo que tratan de decir uno y otro es verdad. El materialismo intenta decir que el mundo consiste por entero de partículas físicas en campos de fuerza. El dualismo intenta decir que el mundo tiene rasgos mentales irreducibles e inerradicables, sobre todo la conciencia y la intencionalidad. Pero si ambas cosmovisiones son verdaderas, debe haber una manera de enunciarlas que las haga compatibles. Dadas las ca-

tegorías tradicionales, no es fácil ver cómo podrían llegar a serlo; pues el materialismo, así enunciado, parece dar a entender que no puede haber ningún fenómeno no físico irreducible; y el dualismo, expuesto según esas categorías, parece sugerir que, por añadidura a los fenómenos materiales, debe haber fenómenos mentales no físicos irreducibles. Exploraremos estas cuestiones con más detalle en el próximo capítulo y veremos que, a fin de hacer congruentes ambas concepciones, es preciso abandonar los supuestos subyacentes al vocabulario tradicional.

CAPÍTULO 4

LA CONCIENCIA, PRIMERA PARTE

LA CONCIENCIA Y EL PROBLEMA MENTE-CUERPO

Terminamos el capítulo anterior con una aparente contradicción, de la clase que es típica en filosofía. Por un lado aceptamos una concepción que parece abrumadoramente convincente –el universo es material–, pero al parecer incompatible con otra perspectiva a la que no podemos renunciar: la mente existe. Este patrón se reitera una y otra vez en la filosofía. En el capítulo 7 veremos que el problema del libre albedrío muestra el mismo tipo de conflicto o contradicción: creemos que todos los sucesos deben tener una determinación causal, pero experimentamos la libertad. En otras ramas de la filosofía surgen inconsistencias similares. En ética sentimos que debe haber una verdad moral objetiva, pero al mismo tiempo nos parece que ese tipo de objetividad no tiene cabida en la moral. Algunas personas consideran exasperantes las contradicciones en filosofía. Otras, como yo, las juzgan divertidas e intrigantes.

En este capítulo voy a tratar de resolver la contradicción entre mente y materia.

I. Cuatro supuestos erróneos

Hasta ahora, en este libro, me he ocupado sobre todo de las opiniones de otras personas. He intentado describir la configuración del terreno, y sólo agregué mi opinión cuando parecía parte de ella. Y utilicé incluso la terminología aceptada, aunque la considero inadecuada. En este capítulo el lector conocerá lo que real-

mente pienso del "problema mente-cuerpo". Como primer paso, quiero sugerir la necesidad de no aceptar la terminología tradicional y los supuestos que la acompañan. Expresiones como "mente" y "cuerpo", "mental" y "material" o "físico", así como "reducción", "causación" e "identidad", tal cual se emplean en las discusiones sobre el problema mente-cuerpo, son el origen de nuestras dificultades y no herramientas para su resolución. Como mi solución al problema mente-cuerpo se contrapone a esos supuestos, quiero exponerlos de manera explícita (con comentarios preliminares entre paréntesis). Cuatro son los supuestos que es preciso cuestionar.

Supuesto 1. La distinción entre lo mental y lo físico

Se supone que "mental" y "físico" se refieren a categorías ontológicas mutuamente excluyentes. Si algo es mental, no puede ser físico en ese mismo aspecto. Y si es físico, no puede ser mental. Lo mental como tal excluye lo físico como tal.

(Este es el supuesto básico, y el que mantiene en marcha todo el debate. Si consideramos que el mundo, en el fondo, es físico, ¿cómo debemos concebir el encaje de lo mental en él? Una actitud habitual de la gente que cree negar este supuesto consiste en decir que podemos *reducir* lo mental a lo físico. Lo mental no es más que lo físico. Estas personas creen superar de un modo u otro la dicotomía dualista, pero en realidad aceptan su peor rasgo. Cuando dicen que lo mental es físico, no dicen que lo mental como tal es físico como tal. Dicen que lo mental como tal no existe: sólo lo físico existe. Este es un punto crucial, al cual volveré más adelante.)

Supuesto 2. La noción de reducción

En general se supone que la noción de reducción, por la cual un tipo de fenómeno se reduce a otro tipo, es clara, inequívoca y no problemática. Cuando reducimos A a B, mostramos que A no es otra cosa que B. Los objetos materiales, por ejemplo, pueden reducirse a moléculas porque no son otra cosa que agrupaciones de estas. De manera análoga, si la conciencia puede reducirse a los procesos cerebrales, significará que no es más que un proceso cerebral.

(El modelo de la reducción procede de las ciencias naturales. Así como la ciencia ha mostrado que los objetos materiales sólo son agrupaciones de moléculas, también podría demostrar que la conciencia no es sino otra cosa: las activaciones de neuronas y los programas informáticos son los candidatos preferidos. Más adelante veremos que esta noción es ambigua en múltiples aspectos. Será preciso trazar una distinción entre las reducciones que eliminan el fenómeno reducido al mostrar que es una ilusión –las puestas de sol, por ejemplo, se eliminan al demostrar que son una ilusión generada por la rotación de la Tierra–, y las que muestran cómo se realiza en el mundo un fenómeno real: los objetos materiales, por ejemplo, se reducen a moléculas, pero eso no significa que no existan. También deberemos distinguir entre reducciones causales y reducciones ontológicas.)

Supuesto 3. Causalidad y sucesos

Se supone de manera casi universal que la causación es siempre una relación entre sucesos discretos ordenados en el tiempo, en los que la causa precede al efecto. Un suceso, la causa, aparece antes que otro, el

efecto. Los casos específicos de relaciones de causa y efecto deben ejemplificar una ley causal universal.

(Como consecuencia inmediata de los supuestos 1 y 3, si los sucesos cerebrales causan sucesos mentales, se sigue que hay dualismo. El suceso cerebral es una cosa [física]. El suceso mental es otra cosa [mental].)

Supuesto 4. La transparencia de la identidad

Se supone que la identidad, como la reducción, no plantea problema alguno. Todo es idéntico a sí mismo y distinto de todo lo demás. Los paradigmas de la identidad son las identidades de objetos y las identidades de composición. Un ejemplo de la primera: el objeto "lucero vespertino" es idéntico al objeto "lucero matutino". Un ejemplo de identidad de composición: el agua es idéntica a las moléculas de H_2O porque cualquier extensión de agua está compuesta de H_2O.

(El motivo de la introducción del concepto de identidad en esta discusión radica en que podríamos descubrir que un estado mental es idéntico a un estado neurofisiológico del cerebro, del mismo modo como hemos descubierto que el lucero vespertino es idéntico al lucero matutino o que el agua es H_2O.)

Creo que estos supuestos contienen tremendas confusiones. Mi método no consistirá en atacarlos de frente; al menos, no lo haré así por el momento. En primer lugar quiero abordar la relación de la conciencia con los procesos cerebrales de una manera ingenua, como si no tuviéramos a nuestras espaldas muchos siglos de confusión motivada. Luego, después de explicar las relaciones de la mente y el cuerpo, volveré atrás y explicaré por qué esos supuestos, tal cual son, nos han impedido hacernos una idea más clara de los hechos y necesitan una seria rectificación y revisión.

II. La solución al problema mente-cuerpo

Mi método en filosofía consiste en tratar de olvidar la historia de un problema y los modos tradicionales de pensarlo, para enunciar sencillamente los hechos hasta donde los conocemos. Probemos el método con un caso bastante simple. Nos concentraremos en la conciencia y abordaremos la intencionalidad en un próximo capítulo. Aquí vamos: en este momento tengo sed. No una sed desesperada, apenas un deseo consciente y moderado de tomar un poco de agua. Esa sensación, como todos los estados conscientes, sólo existe en cuanto es experimentada por un sujeto humano o animal, y en ese sentido tiene una ontología subjetiva o de primera persona. Para existir, las sensaciones como mi sed deben ser vividas por un sujeto, un "yo" ["I"] que está sediento. Pero ¿cómo se ajustan esas sensaciones subjetivas de sed al resto del mundo? Ante todo, es preciso insistir en que mi sed es un fenómeno real, una parte del mundo real, y que actúa causalmente en mi comportamiento. Si ahora bebo, es porque tengo sed. A continuación debemos advertir que mi sensación es íntegramente causada por procesos neurobiológicos con sede en el cerebro. Si no tengo agua suficiente en mi sistema, esa escasez desencadena una compleja serie de fenómenos neurobiológicos, y todos ellos causan mi sensación de sed. (De paso, hay una extraña renuencia a admitir que nuestros estados conscientes son causados por procesos cerebrales. Algunos autores inventan y dicen que el cerebro "da origen" a la conciencia[1];

1 D. Chalmers, *The Conscious Mind: In Search of a Theory of Conscious Experience*, Nueva York, Oxford University Press, 1996, pp. 115-121.

otros dicen que el cerebro es su "sede"[2]. Uno de los que admite que la conciencia depende del cerebro dice que la relación se "concibe poco felizmente como causal"[3].) Pero ¿qué es exactamente esa sensación de sed? ¿Dónde y cómo existe? Es un proceso consciente producido en el cerebro, y en ese sentido es un rasgo de este, aunque en un nivel superior al de las neuronas y sinapsis. La sensación consciente de sed es un proceso en desarrollo dentro de mi sistema cerebral.

Para que no parezca que hablo vagamente de cómo podrían ser las cosas en contraste con su modo de ser en los hechos, permítanme anclar toda la cuestión en la realidad, para lo cual resumiré parte de lo que sabemos acerca del papel de los procesos cerebrales como causa de la sensación de sed. Supongamos que un animal tiene escasez de agua en su sistema. Esa escasez motivará "desequilibrios salinos" en el sistema, porque la proporción entre la sal y el agua es excesiva en beneficio de la primera. La situación desencadena ciertas actividades en los riñones. Estos secretan renina, y la renina sintetiza una sustancia denominada angiotensina 2. Esta sustancia penetra en el hipotálamo y afecta la velocidad de las activaciones neuronales. Por lo que sabemos, las velocidades diferenciales de esas activaciones hacen que el animal sienta sed. Ahora bien, no conocemos todos los detalles, desde luego, y como es de imaginar, cuando lleguemos a saber más este breve esbozo que acabo de dar parecerá bastante

2 T. Huxley, "On the Hypothesis that Animals Are Automata and Its History", en D. M. Armstrong (comp.), *The Mind-Body Problem: An Opinionated Introduction*, Boulder, Westview Press, 1999, p. 148.

3 J. Kim, *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*, Cambridge (Mass.), MIT Press, 1998, p. 44.

peregrino. Pero así se explica cómo encaja la existencia de una sensación consciente de sed en nuestra cosmovisión general. Todas las formas de conciencia son causadas por el comportamiento neuronal y se realizan en el sistema cerebral, compuesto a su vez de neuronas. Lo que vale para la sed vale para todas las formas de nuestra vida consciente, desde la necesidad de vomitar hasta el hecho de preguntarse cómo traducir los poemas de Stéphane Mallarmé a un inglés coloquial. Todos los estados conscientes tienen como causa procesos neuronales de nivel inferior localizados en el cerebro. Tenemos pensamientos y sentimientos conscientes, causados por procesos neurobiológicos en el cerebro; y esos pensamientos y sentimientos existen como características biológicas del sistema cerebral.

En mi opinión, esta sucinta descripción representa un inicio de solución del "problema mente-cuerpo": sospecho de los "ismos", pero a veces es útil contar con un nombre, aunque sólo sea para distinguir con claridad entre una concepción y otra. Doy a mi perspectiva el nombre de "naturalismo biológico" porque aporta una solución naturalista al tradicional "problema mente-cuerpo", una solución que hace hincapié en el carácter biológico de los estados mentales y evita tanto el materialismo como el dualismo.

Expondré el naturalismo biológico con respecto a la conciencia como un conjunto de cuatro tesis:

1. Los estados conscientes, con su ontología subjetiva de primera persona, son fenómenos reales del mundo real. No podemos hacer una reducción eliminativa de la conciencia y mostrar que es una mera ilusión. Tampoco podemos reducirla a sus fundamentos neurobiológicos, porque esa

reducción de tercera persona excluiría su ontología de primera persona.

2. Los estados conscientes son causados en su totalidad por procesos neurobiológicos de nivel inferior con sede en el cerebro. Por ello, son *causalmente reducibles* a procesos neurobiológicos. No tienen absolutamente ninguna vida propia, independiente de la neurobiología. Hablando en términos causales, no están "por encima" de los procesos neurobiológicos.
3. Los estados conscientes se realizan en el cerebro como rasgos del sistema cerebral y existen, por lo tanto, en un nivel superior al de las neuronas y sinapsis. Por sí misma, una neurona no es consciente, pero las partes del sistema cerebral compuestas por ellas sí lo son.
4. Como los estados conscientes son características reales del mundo real, funcionan en forma causal. Mi sed consciente, por ejemplo, me lleva a tomar agua. En el capítulo 7, "La causación mental", explicaré en detalle el funcionamiento de este proceso.

¿Puede ser realmente tan simple la solución al famoso "problema mente-cuerpo"? Si somos capaces de salir de las categorías tradicionales, creo, en efecto, que es muy simple. Sabemos sin duda alguna que todos nuestros procesos mentales son causados por procesos neurobiológicos, y también sabemos que se producen en el cerebro y quizás en el resto del sistema nervioso central. Sabemos que funcionan causalmente, aunque no tienen facultades causales al margen de las correspondientes a la neurobiología subyacente, y sabemos asimismo que no pueden ser objeto de una reducción

ontológica a fenómenos de tercera persona, porque tienen una ontología de primera persona. ¿Por qué, entonces, esta solución aparentemente obvia tropieza con tantas resistencias? Muchos filósofos no comprenden cómo pueden existir estas entidades mentales presuntamente misteriosas, y si existen, cómo pueden ser causadas por procesos físicos en bruto localizados en el cerebro, y si existen y son causadas por procesos físicos, cómo pueden existir en el sistema físico cerebral. Adviértase, empero, que esta forma de plantear las dificultades y los interrogantes ya acepta el dualismo de lo mental y lo físico. Si formulamos la tesis sin recurrir al vocabulario cartesiano tradicional, el misterio se desvanecerá por completo. Mi sensación consciente de sed existe efectivamente y funciona de manera causal en mi comportamiento (¿duda alguien que haya sentido sed alguna vez de su existencia y poder causal?). Sabemos a ciencia cierta que es causada por procesos neuronales, y la sensación misma es un proceso que ocurre dentro del cerebro.

III. La superación de los supuestos erróneos

Para ver por qué es tan difícil aceptar estos argumentos, volvamos un momento atrás y examinemos los cuatro supuestos que, según dije, imposibilitan alcanzar una solución al llamado problema mente-cuerpo.

Supuesto 1. ***La distinción entre lo mental y lo físico***

El peor error consiste en suponer que la distinción de sentido común entre estados mentales y estados físicos, unos y otros interpretados de manera ingenua, es la expresión de una profunda distinción metafísica. De

acuerdo con la concepción que expongo, no lo es. La conciencia es una característica biológica sistémica, como lo son la digestión, el crecimiento o la secreción de bilis. En ese carácter, es un rasgo del cerebro y, con ello, una parte del mundo físico. La tradición contra la cual milito dice que, por ser intrínsecamente mentales, los estados mentales no pueden ser físicos desde ese mismo punto de vista. Por mi parte, digo en sustancia que, como son intrínsecamente mentales, constituyen un tipo determinado de estado biológico y *a fortiori*, por lo tanto, son físicos. Sin embargo, toda la terminología de lo mental y lo físico tiende por naturaleza a plantear una oposición absoluta entre uno y otro término, por lo cual acaso sea mejor no utilizarla y limitarse a decir que la conciencia es un rasgo biológico del cerebro al igual que la digestión es un rasgo biológico del tracto digestivo. En ambos casos hablamos de procesos naturales. No hay abismo metafísico.

El inconveniente que enfrentamos con la terminología es que, por tradición, los términos se han definido como recíprocamente excluyentes. "Mental" se define como cualitativo, subjetivo, de primera persona y por ende inmaterial. "Físico" se define como cuantitativo, objetivo, de tercera persona y por consiguiente material. Sugiero que estas definiciones son inadecuadas para aprehender el hecho de que el mundo funciona de tal manera que algunos procesos biológicos son cualitativos, subjetivos y de primera persona. Si pretendemos mantener la terminología, necesitaremos una noción ampliada de lo físico que permita dar cabida a su componente mental intrínseco y subjetivo. Hagámoslo, pues. Elaboremos una lista de los rasgos tradicionales de lo mental y lo físico que se estiman excluyentes entre

sí y revisémosla hasta donde sea necesario para ajustarla a los hechos.

Según la concepción tradicional, si algo es mental tiene las características listadas en la izquierda; si es físico, las de la derecha[4].

Mental	Físico
Subjetivo	Objetivo
Cualitativo	Cuantitativo
Intencional	No intencional
No situado espacialmente y No extendido en el espacio	Espacialmente situado y Extendido en el espacio
No explicable a través de procesos físicos	Causalmente explicable mediante el recurso a la microfísica
Incapaz de actuar causalmente sobre lo físico	Actúa de manera causal y como sistema está causalmente cerrado

Los rasgos de lo mental que es preciso explicar en una teoría unificada de la totalidad son la conciencia y la intencionalidad. Las características relevantes de la conciencia son su cualitatividad y su subjetividad (juntas, ambas implican la "primera persona", por lo cual no es necesario mencionar esta última como un rasgo especial). El interrogante es: ¿cómo encajan en el mundo físico los fenómenos cualitativos, subjetivos e intencionales? ¿Cuáles son los rasgos del mundo físico a los que deben ajustarse? El concepto contemporáneo de lo

4 Se encontrará una versión anterior de esta lista en H. Feigl, "The 'Mental' and the 'Physical'", *op. cit.*

físico es mucho más complejo de lo admitido por la tradición cartesiana. Por ejemplo, si los electrones son puntos de masa y energía, no son físicos según la definición de Descartes, porque carecen de extensión. Pero cualquier concepción razonable de lo físico exige al menos estas características formales: en primer lugar, los fenómenos físicos reales están situados en el espacio-tiempo. (Así, los electrones son físicos y los números no lo son.) Segundo, sus rasgos y comportamientos pueden explicarse causalmente mediante el recurso a la microfísica. (La solidez y la liquidez satisfacen este criterio. Los fantasmas, si existieran, no lo harían.) Tercero, cuando son reales, los fenómenos físicos funcionan de manera causal. (Así, la solidez es un fenómeno físico real. El arco iris, de acuerdo con la definición "arco iris", no es un arco físico real en el cielo. No causa nada.) Y el universo físico está causalmente cerrado en el sentido trivial de que todo lo que funciona de manera causal en él debe ser parte de él.

Ahora examinemos las listas. Los primeros tres rasgos de la columna mental son perfectamente compatibles con los últimos cuatro de la columna física. Vale decir, la cualitatividad, la subjetividad y la intencionalidad son físicas de acuerdo con los últimos cuatro criterios. Están localizadas en el espacio del cerebro en determinados períodos, son causalmente explicables por medio de procesos de nivel inferior y pueden actuar de manera causal. ¿Qué pasa con los otros rasgos? Los últimos cuatro de la lista mental son sencillamente erróneos. Un fenómeno mental no tiene como condición ser no espacial, no explicable por microprocesos y causalmente inerte. Los primeros tres rasgos tampoco implican estos cuatro. Al contrario, toda mi vida mental ocurre en el espacio del cerebro, donde es causada por

microprocesos, y actúa causalmente desde allí. Bien, ¿y qué pasa entonces con los primeros tres de la lista física? No son condiciones necesarias para formar parte del universo físico. *No hay razón por la cual un sistema físico como un organismo humano o animal no deba tener estados cualitativos, subjetivos e intencionales.* De hecho, en la vida real, los estudios de los sistemas perceptivo y cognitivo son justamente casos de tratamiento de la cualitatividad, la subjetividad y la intencionalidad original como parte del dominio de las ciencias naturales y, por ende, del mundo físico. Digamos de paso que la distinción entre cantidad y calidad es probablemente espuria. No hay razones metafísicas que impidan hacer mediciones de las magnitudes del dolor o la percepción consciente, por ejemplo.

Este es uno de los mensajes más importantes del presente libro. Una vez que revisamos las categorías tradicionales en procura de ajustarlas a los hechos, no hay inconveniente en reconocer que lo mental en cuanto mental es físico en cuanto físico. Es preciso revisar las definiciones cartesianas tradicionales de lo "mental" y lo "físico", pero ninguna de las dos, de todos modos, se ajustaba a los hechos.

Supuesto 2. *Reducción*

Las nociones de reducción y "reducibilidad" [*reducibility*] se cuentan entre las más confusas de la filosofía, porque están plagadas de ambigüedades. En primer lugar debemos distinguir entre reducciones causales y reducciones ontológicas. Podemos decir que los fenómenos de tipo A son *causalmente reducibles* a los fenómenos de tipo B si y sólo si el comportamiento de A es totalmente explicable en términos causales por el comportamiento de B, y A no tiene facultades causales

al margen de las de B. Así, por ejemplo, la solidez es causalmente reducible al comportamiento molecular. Las características de los objetos sólidos –impenetrabilidad, capacidad de soportar otros objetos sólidos, etc.– se explican de manera causal a través del comportamiento molecular, y la solidez no tiene poderes causales adicionales a las facultades correspondientes de las moléculas. Los fenómenos de tipo A son *ontológicamente* reducibles a los fenómenos de tipo B si y sólo si A no es otra cosa que B. Así, por ejemplo, los objetos materiales no son otra cosa que agrupamientos de moléculas, y las puestas de sol no son otra cosa que apariencias generadas por la rotación de la Tierra sobre su eje en relación con el Sol.

En la historia de la ciencia hacemos a menudo –casi siempre, a decir verdad– una *reducción ontológica* sobre la base de una *reducción causal*. Decimos: la solidez no es otra cosa que una clase determinada de comportamiento molecular. Desechamos sus características superficiales, como el hecho de que los objetos sólidos tienen cierto tacto, resisten la presión y son impenetrables por otros objetos, y redefinimos el concepto en función de las causas subyacentes. Ahora, la solidez no se define en términos de las características superficiales sino desde la perspectiva del comportamiento molecular. Y aquí está el quid para nuestro presente análisis: *en el caso de la conciencia podemos hacer una reducción causal, pero no podemos hacer una reducción ontológica sin perder de vista el sentido del concepto*. La conciencia recibe una completa explicación causal a través del comportamiento neuronal, pero con ello no se demuestra que no sea otra cosa que ese comportamiento. ¿No podríamos, acaso, hacer una reducción ontológica y decir que la conciencia no es sino comportamiento neuronal? Bien, sí, podríamos, y con una finalidad médica o algún otro propósito científico sería posible redefinirla en términos de microsustratos, como hemos hecho con la solidez y la liquidez. De ese modo podríamos decir, por ejemplo: "Este tipo tiene efectivamente un dolor, pero todavía no puede sentirlo. Nuestro cerebroscopio muestra la presencia de dolor en el sistema talamocortical". En forma análoga, ahora podemos decir: "El vidrio es realmente líquido, aunque a la vista y al tacto parezca sólido". Pero el principal sentido del concepto de conciencia es la posibilidad de aprehender los rasgos subjetivos y de primera persona del fenómeno, y ese sentido se pierde si redefinimos aquella en términos objetivos y de tercera persona. De hacerlo, seguiremos necesitando un nombre para la ontología de primera persona. La conciencia, entonces, difiere de otros fenómenos como la liquidez y la solidez que tienen características superficiales, en cuanto somos reacios a desechar estas últimas y redefinir la noción desde el punto de vista de las causas que las motivan, porque el sentido del concepto es identificarlas. Hay muchos conceptos en los que las características superficiales de los fenómenos son más interesantes que la microestructura. Consideremos el lodo o la novena sinfonía de Beethoven. El comportamiento del lodo es comportamiento molecular, pero lo interesante no es eso, de modo que pocas personas se afanan en decir: "El lodo puede reducirse al comportamiento molecular", aunque podrían hacerlo si realmente quisieran. Otro tanto ocurre con Beethoven. Las interpretaciones de la novena sinfonía pueden reducirse a movimientos de ondas en el aire, pero no eso lo que nos interesa en la ejecución. El crítico musical que escribiera: "Sólo pude oír movimientos de ondas", habría pasado por alto el sen-

tido de la interpretación. Podríamos hacer, de manera análoga, una reducción de la conciencia y la intencionalidad, pero de todos modos necesitaríamos un vocabulario para hablar de las características superficiales. La singularidad de la conciencia y la intencionalidad sólo reside en el hecho de tener una ontología de primera persona.

En una versión anterior de este argumento (*The Rediscovery of the Mind*) dije que la irreductibilidad de la conciencia era una consecuencia trivial de nuestras prácticas definicionales. Como la observación fue objeto de un malentendido generalizado, permítanme aclararla aquí. Debe concedérseme que el mundo "físico" real contiene entidades con una ontología de tercera persona (árboles y hongos, por ejemplo) y entidades con una ontología de primera persona (dolores y experiencias del color, por ejemplo). Todas estas entidades de primera persona son causalmente reducibles a sus fundamentos causales de tercera persona. Pero hay una asimetría. Cuando se trata del color estamos dispuestos (o al menos algunos lo estamos) a desechar las experiencias conscientes, las experiencias del color con su ontología de primera persona, a hacerlas a un lado para redefinir luego las palabras relacionadas con los colores en términos de tercera persona. Según una concepción, los colores no se definen en esencia desde el punto de vista de las experiencias suscitadas por ellos, sino en términos de la reflexión de la luz causante de dicha experiencia. Sin embargo, no estamos dispuestos a hacer lo mismo en el caso de la conciencia y de conceptos correspondientes a ella como el dolor. ¿Por qué no? ¿Por qué no extraemos las experiencias de primera persona de la conciencia y el dolor, las hacemos a un

lado y redefinimos los conceptos en función de sus causas, como hicimos con el color?

Bueno, podríamos hacerlo, y si supiéramos mucho más acerca de las causas, en ciertos aspectos quizá lo haríamos. Pero hay una asimetría entre los colores, por un lado, y los dolores y la conciencia, por otro, porque perderíamos de vista el sentido de los conceptos de la conciencia si desecháramos la ontología de primera persona y redefiniéramos las palabras en términos de tercera persona. En ese aspecto, la irreductibilidad de la conciencia no revela una profunda asimetría metafísica entre, por ejemplo, la relación de las experiencias del color con sus causas y la relación de las experiencias dolorosas con las suyas; la asimetría está, por el contrario, en nuestras prácticas definicionales. Pues para definir el "dolor" las sensaciones generadas por él nos preocupan más que las producidas por el "color" cuando, a su turno, debemos definirlo.

Algunos de mis críticos vieron en mi postura la intención de afirmar que la existencia misma de la conciencia era una consecuencia trivial de nuestras prácticas definicionales. Sin embargo, no digo nada de eso. Espero que esto aclare el malentendido.

Pero ¿acaso las reducciones no se libran del fenómeno reducido al mostrar que es realmente otra cosa? No, y esto conduce a la segunda confusión en el concepto de reducción. Es menester distinguir entre las reducciones que son *eliminativas* y las que no lo son. Las primeras muestran que el fenómeno reducido en realidad no existía. Así, la reducción de las puestas de sol a la rotación de la Tierra es eliminativa porque demuestra que aquellas son una simple apariencia. Pero la reducción de la solidez no es eliminativa de ese modo,

porque no muestra, por ejemplo, que los objetos no oponen una resistencia real a otros objetos. No se puede hacer una reducción eliminativa de algo que tiene existencia real.

¿Por qué no podríamos mostrar, empero, que la conciencia es una ilusión como las puestas de sol y hacer así una reducción eliminativa? Las reducciones de este tipo se apoyan en la distinción entre apariencia y realidad. Pero no podemos mostrar que la existencia misma de la conciencia es una ilusión como las puestas de sol, porque en lo que a ella respecta la apariencia es la realidad. El Sol parece ponerse sobre el monte Tamalpais, aunque en realidad no es así. Pero si conscientemente me parece que soy consciente, entonces lo soy. Puedo cometer toda clase de errores acerca de los contenidos de mis estados conscientes, pero no con respecto a su existencia misma.

Resumamos esta breve discusión sobre la reducción: no se puede hacer una reducción eliminativa de la conciencia porque esta existe realmente; y su existencia real no está sujeta a las dudas epistémicas usuales, basadas en una distinción entre apariencia y realidad que es imposible hacer en el caso de la existencia de nuestros propios estados conscientes. Podemos hacer una reducción causal de la conciencia a su sustrato neuronal, pero no llegamos así a una reducción ontológica, porque la conciencia tiene una ontología de primera persona y el sentido del concepto se desvanece si lo redefinimos en términos de tercera persona.

Supuesto 3. *Causalidad y sucesos*

Muchas relaciones causales se dan entre sucesos discretos ordenados en el tiempo. Un caso paradigmático, muy apreciado por los filósofos, es el de la bola de billar que golpea otra y se detiene, mientras la segunda se aleja. Pero la causación no siempre es así. En muchos casos la causa es simultánea con el efecto. Si el lector observa los objetos que lo rodean, notará que ejercen presión sobre el suelo de la habitación donde se encuentran. ¿Cuál es la explicación causal de esa presión? La fuerza de gravedad. Pero esta no es un suceso separado. Es una fuerza continua que actúa en la naturaleza. Por otra parte, muchos casos de causación simultánea se dan, por así decirlo, de abajo arriba, en el sentido de que microfenómenos de un nivel inferior causan macrorrasgos de un nivel superior. Vuelva el lector a mirar los objetos a su alrededor. La mesa sostiene libros. El hecho de que los sostenga se explica causalmente por el comportamiento de las moléculas. Para el caso de la solidez, como ya he mencionado, hacemos una reducción ontológica sobre la base de la reducción causal. Pero la terminología podría haber seguido uno u otro camino. Podríamos haber dicho que la solidez tiene que ver con la resistencia de las cosas a la presión, su impenetrabilidad y su capacidad de servir de apoyo a otros objetos. Y esto se explica en términos causales por el comportamiento de las moléculas. No elegimos ese camino porque a nuestro juicio la microestructura nos brinda una explicación más profunda. Decimos que la solidez es simplemente el movimiento vibratorio de las moléculas en estructuras reticuladas, y eso explica el hecho de que un objeto sostenga otro. El quid, sin embargo, es que examinamos el orden causal de la naturaleza, y ese orden no suele ser cuestión de sucesos discretos y secuenciales en el tiempo, sino de microfenómenos que explican causalmente macrorrasgos de sistemas.

Supuesto 4. *Identidad*

Los criterios de identidad para objetos materiales como los planetas y tipos de compuestos como el agua son razonablemente claros. Pero cuando se trata de sucesos, como la Gran Depresión o mi fiesta de cumpleaños, los criterios no son tan claros. Al considerar los sucesos mentales, como el hecho de que yo tenga cierta experiencia, debemos decidir la magnitud que queremos asignarles. ¿La conciencia es idéntica a un proceso cerebral o no? Bueno, desde un punto de vista obvio y trivial, como he dicho, la conciencia es sólo un proceso cerebral. Es un proceso cualitativo, subjetivo y de primera persona que ocurre en el sistema nervioso. Sí, pero no es eso lo que querían los teóricos de la identidad. Su ambición era identificar un estado consciente con un proceso neurobiológico, descrito en términos neurobiológicos. Me parece que aquí buscamos una decisión y no un descubrimiento. Creo que podemos considerar que un mismo suceso tiene rasgos neurobiológicos y rasgos fenomenológicos. Un mismo suceso es una secuencia de activaciones neuronales y a la vez provoca dolor. Pero este tipo de identidad no aporta a los materialistas lo que estos quieren. La cuestión se asemeja un poco al ejemplo de las identidades de casos propuesto por Jaegwon Kim[5]. Cada objeto coloreado específico es idéntico a un objeto específico con forma. Eso es indudablemente cierto, pero no muestra que el hecho de tener color y el hecho de tener forma sean lo mismo. De manera semejante, podemos contar con una noción de los procesos neurobiológicos de magnitud

5 J. Kim, *The Philosophy of Mind*, Boulder, Westview Press, 1998, p. 59.

suficiente para que cada proceso doloroso específico sea un proceso neurobiológico específico en el cerebro, pero de ello no se deduce que la sensación de dolor de primera persona sea igual al proceso neurobiológico de tercera persona. El concepto de identidad no nos sirve de mucho en el problema mente-cuerpo, porque podemos hacer que nuestros sucesos sean lo bastante grandes para incluir tanto el aspecto fenomenológico como el aspecto neurobiológico. Lo atinado, como siempre, es olvidar esas grandes categorías y tratar de describir los hechos, para luego volver y ver cómo se deben ajustar los preconceptos que uno tenga de las otras categorías a fin de dar cabida a esos hechos.

Pero si de la definición de nuestro suceso se desprende que este tiene rasgos fenomenológicos y neurobiológicos, ¿no estará la identidad resultante sujeta a la objeción de Kripke sobre las identidades necesarias? No. En el caso de la identidad necesaria entre el agua y el H_2O, la necesidad se alcanza a través de la redefinición. Una vez que descubrimos que la sustancia que hemos llamado agua está compuesta de moléculas de H_2O, incluimos "H_2O" en la definición de "agua". Que el agua es H_2O se convierte entonces en una verdad necesaria. De manera similar, podemos reajustar nuestras definiciones a fin de que parte de lo que hace de este dolor el dolor que es sea el hecho de ser causado por este tipo de proceso neurobiológico y se realice en él. Parte de lo que hace de ese proceso neurobiológico el proceso que es, es el hecho de causar y realizar aquel mismo dolor. Dicho sea de paso, la definición de las sensaciones en términos de sus causas es muy común. Considérese "ciática", definida como un tipo de dolor causado por la estimulación del nervio ciático.

IV. Ni materialismo ni dualismo

Vale la pena destacar que la concepción que expongo difiere tanto del materialismo como del dualismo. Como creo que uno y otro tratan de decir algo cierto, es importante separar en ambos las partes verdaderas de las partes falsas. Para hacerlo necesito enunciar con exactitud las diferencias entre mi punto de vista y esas doctrinas tradicionales. El materialismo intenta decir con veracidad que el universo está íntegramente constituido de partículas físicas existentes en campos de fuerza y a menudo organizadas en sistemas. Pero termina por incurrir en una falsedad al decir que no hay fenómenos mentales ontológicamente irreducibles. El dualismo trata de decir con veracidad que hay fenómenos mentales irreducibles. Pero también termina por caer en una falsedad cuando dice que esos fenómenos están al margen del mundo físico corriente en el que todos vivimos, que son algo situado por encima de su sustrato físico. El desafío consiste en enunciar la parte verdadera de cada concepción y negar la parte falsa. Si nos aferramos al vocabulario tradicional parece imposible hacerlo, porque terminamos por decir que lo mental irreducible (subjetivo, cualitativo) es sólo una parte habitual del mundo físico, lo cual parece autocontradictorio. De modo que, en definitiva, me decido por cuestionar el vocabulario tradicional.

Adviértase que si trato de enunciar mi posición en dicho vocabulario, las palabras significan a la larga algo completamente diferente de su definición según la tradición. El materialista dice: "La conciencia es sólo un proceso cerebral". Yo digo: "La conciencia es sólo un proceso cerebral". Pero el materialista quiere decir: la conciencia como fenómeno irreductiblemente cualitativo,

subjetivo, de primera persona, etéreo y delicado en realidad no existe. Sólo existen los fenómenos objetivos de tercera persona. Por mi parte, yo quiero decir que la conciencia, sin duda como fenómeno irreductiblemente cualitativo, subjetivo, de primera persona, etéreo y delicado, es un proceso que se desenvuelve en el cerebro. El dualista dice: "La conciencia es irreducible a los procesos neurobiológicos de tercera persona". Yo digo: "La conciencia es irreducible a los procesos neurobiológicos de tercera persona". Pero el dualista cree dar a entender con ello que la conciencia no es parte del mundo físico común y corriente, sino algo existente por encima de él. Yo quiero decir que la conciencia es reducible en términos causales, pero no ontológicos. Forma parte del mundo físico común y corriente y no está por encima de él.

Centremos ahora la puntería justamente en esa característica del dualismo. De acuerdo con la concepción de los dualistas, la conciencia es decididamente algo situado por encima de su sustrato material. En rigor, los dualistas suponen que su carácter irreducible ya implica que la conciencia está por encima de su base neurobiológica. Yo rechazo esa implicación. Este punto es tan crucial para todo el argumento del presente libro que voy a exponerlo con cierto detalle. El hecho de que los poderes causales de la conciencia y los poderes causales de su base neuronal sean exactamente los mismos muestra que no hablamos de dos cosas independientes, la conciencia y los procesos neuronales. Si dos cosas pertenecientes al mundo empírico real tienen existencia independiente, deben tener diferentes poderes causales. Pero los poderes causales de la conciencia son exactamente los mismos del sustrato neuronal. Sucede absolutamente lo mismo con los poderes causales de los

objetos sólidos y los poderes causales de sus constituyentes moleculares. No hablamos de dos entidades diferentes sino del mismo sistema en distintos niveles. La conciencia difiere de la solidez, la liquidez, etc., en cuanto la reducción causal no conduce a una reducción ontológica. Y, como hemos visto, sucede así por una razón obvia y hasta trivial. La conciencia tiene una ontología de primera persona; los procesos neuronales tienen una ontología de tercera persona. Por ese motivo, no se puede reducir ontológicamente la primera a los segundos. De tal modo, la conciencia es un aspecto del cerebro, el aspecto consistente en experiencias ontológicamente subjetivas. Pero no hay en nuestro cráneo dos reinos metafísicos diferentes, uno "físico" y otro "mental". Antes bien, sólo hay procesos que se desarrollan en el cerebro, y algunos de ellos son experiencias conscientes.

Dije en el capítulo 3 que los dualistas creen estar en posesión de una intuición profunda que justifica su dualismo. Es hora de dar una respuesta a esa pretensión. La intuición es la siguiente: debe haber una distinción entre lo mental y lo físico, porque una vez fijadas la existencia y las trayectorias de todas las micropartículas del universo, la historia física de este queda determinada en su totalidad por el comportamiento de dichas micropartículas. Sin embargo, cabe concebir aún que no haya estados conscientes en absoluto. Vale decir: desde una perspectiva lógica sería posible que el universo físico fuera exactamente como es, átomo por átomo, pero sin conciencia. Pero, de hecho, no es lógicamente posible que sea tal como es, átomo por átomo, sin que sus características físicas sean exactamente como son. Nótese que este argumento es una ampliación del argumento de los zombis que presenté contra el materialismo en el capítulo 3.

El argumento acierta al señalar que una descripción de los hechos de tercera persona no entraña la existencia de los hechos de primera persona, y ello por la trivial razón de que la ontología de primera persona no puede reducirse a la ontología de tercera persona. Pero el dualista pretende concluir entonces que la conciencia está en otro reino ontológico y es algo situado por encima del cerebro. La conclusión, sin embargo, no se deduce de sus premisas. El dualista deja al margen de este experimento de pensamiento las leyes de la naturaleza. Cuando imaginábamos la trayectoria de las micropartículas, sosteníamos la constancia de todas las leyes naturales. Pero si tratamos de imaginar que esa trayectoria es la misma menos la conciencia, hacemos trampa en el experimento, porque suponemos que las micropartículas *no* se comportan precisamente de la manera como se habrían comportado de actuar de conformidad con aquellas leyes, esto es, en forma tal de causar y realizar estados conscientes (subjetivos y de primera persona). Una vez incluidas las leyes de la naturaleza en la descripción del universo físico –y es preciso incluirlas, porque son parte constitutiva de este–, se sigue la existencia de la conciencia, como consecuencia lógica de esas leyes.

Que un estado de cosas sea o no lógicamente posible depende del modo de describirlo. ¿Es lógicamente posible que haya partículas físicas sin ninguna conciencia en el universo? La respuesta es sí. Sin embargo, ¿es posible que las trayectorias de las partículas físicas existan tal como existieron de hecho junto con las leyes de la naturaleza –que, entre muchas otras cosas, determi-

nan que esas trayectorias causen y realicen la conciencia–, pero sin conciencia alguna? En ese caso la respuesta es no. Descrita de una manera, la ausencia de conciencia es lógicamente posible; descrita de otra manera no lo es. Los dualistas se forjan una imagen en la cual las partículas microfísicas son como diminutos granos de arena afectados por fuerzas independientes, y pueden imaginar el movimiento de la arena sin ninguna conciencia. La imagen, empero, es falsa. En el nivel más fundamental, los puntos de masa y energía están constituidos por las fuerzas descritas por las leyes de la naturaleza. La existencia de la conciencia se deduce de esas leyes como una consecuencia lógica, así como lo hace la existencia de cualquier otro fenómeno biológico, por ejemplo el crecimiento, la digestión o la reproducción.

Una vez más, me parece que la ilusión del dualismo es el producto de la mala comprensión de una distinción muy real. Existe, en efecto, una distinción entre los rasgos irreducibles del mundo que tienen una ontología subjetiva o de primera persona y los que no la tienen. Pero es un profundo error suponer que esa distinción real equivale a la antigua diferencia entre lo mental y lo físico, entre *res cogitans* y *res extensa*, o que los fenómenos subjetivos están por encima de los sistemas en los cuales se realizan.

El dualista cree que la "irreductibilidad" ya implica que el fenómeno irreducible es algo situado por encima de su fundamento físico. Esto plantea un problema imposible al dualista de las propiedades: o bien la conciencia actúa causalmente o bien no lo hace. Si lo hace, tenemos al parecer una sobredeterminación causal: si levanto adrede el brazo, el gesto aparenta tener dos causas, una física y otra mental. Pero si la conciencia

no funciona causalmente, nos topamos con el epifenomenalismo. Ningún problema semejante se presenta para el naturalismo biológico, porque el funcionamiento causal de la conciencia es una forma más del funcionamiento cerebral descrito en un nivel más elevado que el de las neuronas y sinapsis. Piénseselo de esta manera: en términos generales, la conciencia es a las neuronas lo que la solidez del pistón es a las moléculas metálicas. Tanto la conciencia como la solidez funcionan en forma causal. Pero ni una ni otra están "por encima" de los sistemas a los que pertenecen.

V. Resumen de la refutación del materialismo y el dualismo

En el capítulo 3 prometí una refutación del dualismo. En interés de la imparcialidad, agreguemos un enunciado esencial de la refutación del materialismo.

Definamos el materialismo como la concepción de que en el universo no hay otra cosa que fenómenos materiales, según se los concibe tradicionalmente. No hay estados de conciencia intrínsecos y subjetivos irreducibles, ni ninguna otra cosa que sea inherentemente mental. Todo caso aparente puede ser eliminado o reducido a algo físico.

Esta concepción es bastante fácil de refutar, porque niega que existan cosas cuya existencia todos conocemos. Asevera que no hay fenómenos ontológicamente subjetivos, y sabemos que esto es falso porque los experimentamos todo el tiempo. Como filósofos consideramos insatisfactorio este tipo de refutación por su excesiva simpleza, de modo que inventamos argumentos más complejos para plantear la misma cuestión, sobre murciélagos, colores, espectros invertidos, *qualia*,

habitaciones chinas, etc. Pero, cada uno a su manera, todos esos argumentos subrayan el mismo punto.

La refutación del dualismo es más ardua. Definamos esta doctrina como la concepción de que en el universo hay dos reinos metafísicos ontológicamente distintos, uno mental y otro físico. Definición más difícil de refutar, pues mientras el materialismo postulaba la inexistencia de algo cuya existencia todos conocemos, el dualismo postula la existencia de algo, y para refutarlo formalmente habría que probar una negativa universal. En vez de proponer una "refutación" formal, presentaré los argumentos que a mi juicio son concluyentes contra el dualismo.

1. Nadie ha logrado proporcionar jamás una descripción inteligible de las relaciones entre esos dos reinos.
2. La postulación es innecesaria. Es posible explicar todos los hechos de primera persona y todos los hechos de tercera persona sin postular reinos separados.
3. La postulación genera dificultades intolerables. De acuerdo con esta concepción, se hace imposible explicar de qué manera los estados y sucesos mentales pueden causar estados y sucesos físicos. En síntesis, es imposible evitar el epifenomenalismo.

Nótese que estos argumentos no excluyen la posibilidad lógica del dualismo. Es una posibilidad lógica, aunque me parece extremadamente improbable, que, tras la destrucción de nuestros cuerpos, nuestras almas sigan marchando. No he intentado mostrar que es una

imposibilidad (a decir verdad, ojalá fuese cierta), sino que es incompatible con prácticamente todo lo demás que sabemos del funcionamiento del universo, y por lo tanto es irracional creer en ella.

CAPÍTULO 5

LA CONCIENCIA, SEGUNDA PARTE

LA ESTRUCTURA DE LA CONCIENCIA Y LA NEUROBIOLOGÍA

En el capítulo anterior describí cierta ontología básica. Es preciso tenerla presente, con toda su simplicidad y hasta su crudeza, mientras exploramos ahora la notable complejidad y singularidad de la conciencia. Aunque la ontología básica es simple, los fenómenos resultantes son complicados, y los pormenores de sus relaciones neurobiológicas con el cerebro son difíciles de entender y hasta el momento desconocidos en gran parte. Una vez resuelto el problema filosófico, relativamente sencillo, nos quedan por delante problemas neurobiológicos muy arduos.

En este capítulo describiré en primer término la estructura de la conciencia, luego presentaré explicaciones que discrepan de la mía y concluiré con la discusión de algunos de los problemas neurobiológicos de la conciencia.

I. Características de la conciencia

¿Cuáles son las características de la conciencia que cualquier teoría filosófico-científica debe aspirar a explicar? Creo que la mejor manera de proceder consiste en limitarme a enumerar varios de los rasgos centrales de la conciencia humana y presuntamente animal. Aquí van.

1. *Cualitatividad*

Como señalé en capítulos anteriores, todo estado consciente tiene un cariz cualitativo. En ese sentido, los estados conscientes siempre son cualitativos. Dije que algunos filósofos utilizan la palabra "*qualia*" para describir este rasgo, pero a mi entender el término es engañoso en el mejor de los casos, porque su uso sugiere que ciertos estados conscientes no son cualitativos. Al parecer, la idea es que algunos de dichos estados, como el sentir un dolor o el saborear un helado, son cualitativos, pero otros, como la reflexión sobre problemas aritméticos, no tienen un cariz cualitativo especial. Creo que esto es un error. Si el lector supone que no hay cariz cualitativo alguno en pensar que dos más dos es cuatro, trate de pensarlo en francés o en alemán. Para mí es completamente diferente pensar "*zwei und zwei sind vier*", aunque el contenido intencional sea el mismo en alemán y en inglés. Como la noción de conciencia y la noción de *qualia* son totalmente coextensivas, no utilizaré la segunda como algo distinto de la primera y me limitaré a suponer que cuando digo "conciencia", el lector sabe que examino problemas que tienen ese carácter cualitativo.

2. *Subjetividad*

Debido al carácter cualitativo de la conciencia, los estados conscientes sólo existen cuando un sujeto humano o animal los experimenta. Tienen un tipo de subjetividad que yo llamo subjetividad ontológica. Para expresar de otra manera la misma observación, podemos decir que la conciencia tiene una ontología de primera persona. Sólo existe en cuanto un sujeto humano

o animal la experimenta, y en ese sentido sólo existe desde un punto de vista de primera persona. Cuando sé de tu conciencia, tengo un conocimiento que es muy diferente del que tengo de la mía propia.

El hecho de que los estados conscientes sean ontológicamente subjetivos, en el sentido de que sólo existen cuando un sujeto humano o animal los experimenta, no implica que no se los pueda someter a un estudio científico objetivo. Los términos "objetivo" y "subjetivo" oscilan de manera sistemática entre un sentido ontológico y un sentido epistémico. En este último se traza una distinción entre las proposiciones cuya verdad o falsedad puede afirmarse con prescindencia de los sentimientos y actitudes de los hablantes u oyentes, y aquellas en las cuales la verdad o falsedad depende de esos mismos sentimientos y actitudes. Así, el enunciado "Jones mide un metro ochenta centímetros" es epistémicamente objetivo porque su verdad o falsedad no tiene nada que ver con los sentimientos y actitudes del hablante u oyente. El enunciado "Jones es más agradable que Smith", en cambio, es epistémicamente subjetivo porque su verdad o falsedad no puede establecerse con independencia de los sentimientos y actitudes de los participantes en la conversación. Además de este sentido epistémico, hay una distinción entre dos modos de existencia. Los estados conscientes tienen un modo subjetivo de existencia, en cuanto sólo existen cuando son experimentados por un sujeto humano o animal. En este aspecto, difieren de casi todo el resto del universo, por ejemplo las montañas, las moléculas y las placas tectónicas, que tienen un modo objetivo de existencia. El modo de existencia de los estados conscientes es, en efecto, ontológicamente subjetivo, pero la *subjetividad ontológica del tema no impide hacer de él una ciencia*

epistémicamente objetiva. En rigor, toda la ciencia de la neurología exige buscar una descripción científica epistémicamente objetiva de dolores, angustias y otras aflicciones sufridas por los pacientes, a fin de poder tratarlas con técnicas médicas. Cada vez que escucho a filósofos y neurobiólogos decir que la ciencia no puede ocuparse de las experiencias subjetivas, procuro mostrarles libros de texto de neurología en los cuales los científicos y médicos que los escriben, así como quienes los utilizan, no tienen otra alternativa que tratar de proponer una descripción científica de los sentimientos subjetivos de la gente, porque su ambición es ayudar a pacientes reales a aliviar su sufrimiento[1].

3. *Unidad*

En este momento, no sólo experimento sensaciones en la punta de los dedos, la presión de la camisa contra el cuello y la vista de las hojas otoñales mientras caen afuera, sino que vivo todo ello como parte de un solo campo consciente unificado. La conciencia normal y no patológica se nos presenta con una estructura unificada. Kant denominaba "unidad trascendental de la apercepción" esa unidad del campo consciente, y le asignaba mucha importancia. Tenía razón. Como veremos, es inmensamente importante.

En una época yo creía que estos tres rasgos: cualitatividad, subjetividad y unidad, podían describirse

1 Véase, por ejemplo, el capítulo 5, sobre el dolor y la temperatura, de C. R. Noback y R. J. Demarest, *The Nervous System: Introduction and Review*, Nueva York, McGraw-Hill, 1977 [traducción española: *El sistema nervioso: introducción y repaso*, México, Interamericana/ McGraw-Hill, 1993].

como características distintas de la conciencia. Hoy me parece que eso es un error; son aspectos del mismo fenómeno. En su esencia misma, la conciencia es cualitativa, subjetiva y unificada. Es imposible que un estado sea cualitativo, en el sentido al que he hecho referencia, sin ser también subjetivo en el sentido ya explicado. Pero tampoco puede ser cualitativo y subjetivo sin tener el tipo de unidad que acabo de describir. Podremos ver este último punto si tratamos de imaginar nuestro estado actual de conciencia descompuesto en 17 fragmentos independientes. Si eso ocurriera, no tendríamos un estado consciente con 17 partes; habría, antes bien, 17 conciencias independientes, 17 sitios diferentes de la conciencia. Es absolutamente esencial entender que la conciencia no es divisible como suelen serlo los objetos físicos; siempre se presenta en unidades discretas de campos conscientes unificados.

Los llamados experimentos de desconexión callosa o cerebro dividido proporcionan una buena ilustración de este rasgo de la unidad, por lo cual haré una breve digresión para describirlos. Una de las maneras de estudiar la conciencia consiste en estudiar sus formas patológicas o degeneradas, método que utilizaré en diversas oportunidades a lo largo del libro. Los pacientes afectados por el síndrome de desconexión callosa sufrían terribles formas de epilepsia que no podían tratarse mediante los procedimientos normales. Desesperados, los médicos cortaron el cuerpo calloso, la masa de tejido que conecta los dos hemisferios cerebrales. La operación curó de hecho a muchos pacientes epilépticos, pero tuvo otros efectos interesantes. La consecuencia más sorprendente fue que llevó a algunos de ellos a comportarse en ciertas circunstancias como si tuvieran dos centros de conciencia independientes. En un

experimento típico ocurre lo siguiente: se muestra al paciente una cuchara, pero se la coloca en una parte de su campo visual izquierdo, de modo que el estímulo visual sólo vaya al hemisferio derecho de su cerebro. El lenguaje se localiza en el hemisferio izquierdo. Se pregunta entonces al paciente: "¿Qué ve?" Como no tienen percepción visual de la cuchara en el lado izquierdo del cerebro, donde reside el lenguaje, y además, a causa de la escisión del cuerpo calloso, sólo hay una comunicación muy imperfecta entre ambos hemisferios, el paciente responde: "No veo nada". Sin embargo, luego extiende la mano izquierda, controlada por su hemisferio derecho, donde se produce la experiencia visual de la cuchara, y logra tomar el utensilio. Roger Sperry y Michael Gazzaniga realizaron muchos experimentos de este tipo[2]. ¿Tiene el paciente uno o dos centros de conciencia? Por el momento no lo sabemos con total certeza. Pero debemos contemplar al menos la posibilidad de que haya, en efecto, dos campos conscientes dentro de un cerebro, cada uno de ellos correspondiente a un hemisferio, y que en el caso normal ambos se reúnan en un solo campo consciente unificado.

4. *Intencionalidad*

He hablado de la intencionalidad y la conciencia como si fueran fenómenos independientes, pero, desde luego, muchos estados conscientes son intrínsecamente intencionales. Mi presente percepción visual, por ejemplo, no podría ser la experiencia visual que es si

2 M. Gazzaniga, *The Social Brain: Discovering the Networks of the Mind*, Nueva York, Basic Books, 1985 [traducción española: *El cerebro social*, Madrid, Alianza, 1993].

no me pareciera ver sillas y mesas en mi proximidad inmediata. Este rasgo, según el cual muchas de mis experiencias parecen referirse a cosas más allá de sí mismas, es el aspecto que los filósofos han llegado a denominar "intencionalidad". No toda la conciencia es intencional y no toda la intencionalidad es consciente, pero hay muy serias e importantes superposiciones entre una y otra; más adelante veremos que, en realidad, hay conexiones lógicas entre las dos: los estados mentales que son de hecho inconscientes deben ser el tipo de cosa que, en principio, podría convertirse en consciente. Por una serie de razones que van desde el daño cerebral hasta la represión psicológica, pueden ser inaccesibles a la conciencia, pero es preciso que sean la clase de cosa que podría formar parte de un estado mental consciente. Un ejemplo de estado consciente que no es intencional es la sensación de angustia que a veces nos afecta sin un motivo específico; sólo nos sentimos angustiados. Los ejemplos de estados intencionales que no son conscientes son demasiado abundantes para mencionarlos, pero entre los casos evidentes se cuenta el del sueño profundo. Cuando estoy dormido, sigue siendo valedero decir que creo que Bush es presidente y que dos más dos es igual a cuatro, y lo mismo con una gran cantidad de otras creencias que en ese preciso momento no están presentes en mi conciencia.

5. *Humor*

Todos mis estados conscientes se me presentan con un humor u otro. Siempre tengo algún tipo de humor, aunque este carezca de un nombre específico. No hace falta que esté especialmente entusiasmado ni deprimido, y ni siquiera sin ganas de nada; de todas

maneras, hay lo que podríamos llamar cierto sabor en la conciencia, cierto tono en las experiencias conscientes. Un modo de advertirlo es observar los cambios dramáticos. Si recibimos de improviso alguna noticia muy mala, comprobaremos que nuestro humor cambia. Si la noticia es buena, el cambio se dará en la dirección opuesta. El humor no es lo mismo que la emoción porque, en primer lugar, las emociones siempre son intencionales. Siempre tienen algún contenido emocional, mientras que no es imprescindible que el humor lo tenga. Pero los humores nos predisponen a las emociones. Si estamos de humor irritable, es más probable, por ejemplo, que experimentemos la emoción de la ira.

Los humores parecen más susceptibles al control farmacológico artificial que la mayoría de los otros aspectos de la conciencia. Como los dolores, que podemos controlar mediante anestésicos y analgésicos, estamos en condiciones de afectar humores como la depresión por medio de medicamentos como el Prozac y el litio. No es improbable que los avances farmacológicos nos permitan alcanzar un control terapéutico aún más grande de los humores discapacitantes, tal como hicimos con los dolores.

6. La distinción entre el centro y la periferia

Dentro del campo consciente, uno siempre pone más atención a unas cosas que a otras. En este mismo instante me concentro en poner por escrito ciertas ideas sobre la filosofía de la mente, y no en los sonidos procedentes del exterior o la luz que entra en abundancia por la ventana. Algunas cosas están en el centro de mi campo consciente, y otras en la periferia. Un buen indicio de ello es la capacidad de reorientar la atención a

voluntad. Puedo centrarla en el vaso de agua frente a mí o en los árboles que veo por la ventana sin modificar siquiera la postura ni mover los ojos. En cierto sentido, el campo consciente sigue siendo el mismo, pero enfoco algunos de sus rasgos y no otros. Esta aptitud de reorientar la atención y la distinción entre los rasgos del campo consciente que tenemos y no tenemos en cuenta ya es un tema de investigación importante en neurobiología.

Por añadidura a nuestra capacidad de desplazar la atención a voluntad, el cerebro suele hacer pequeños trucos para compensar ciertas deficiencias. No vemos nuestro punto ciego, aunque lo tenemos, y vemos el color en la periferia de nuestro campo visual aun cuando en ella no hay receptividad a los colores.

7. Placer/displacer

En conexión con el humor, pero no idéntico a él, debemos señalar el fenómeno por el cual cada estado consciente despierta cierto grado de placer o displacer. O bien cabría decir, mejor, que se sitúa en alguna posición dentro de una escala que incluye las nociones corrientes de placer y displacer. Así, con respecto a cualquiera de nuestras experiencias conscientes, es legítimo preguntar: ¿la disfruté? ¿Fue divertida? ¿La pasé bien, mal, me aburrí, me entretuve? ¿Fue repugnante, deliciosa o deprimente? Cuando se trata de la conciencia, la dimensión del placer y el displacer es ubicua.

8. Situacionalidad

Todas nuestras experiencias conscientes están acompañadas por una sensación de lo que podríamos

llamar la situación contextual en la que experimentamos el campo de la conciencia. Esa sensación de la situación no debe ser por fuerza parte del campo consciente, y en general no lo es. Pero por lo común sé, en algún sentido, en qué lugar de la superficie de la tierra me encuentro, qué hora es, en qué época del año estamos, si he almorzado o no, de qué país soy ciudadano, etc., dentro de una gama de características que doy por sentadas como la situación correspondiente a mi campo consciente. Uno cobra conciencia de la sensación de situacionalidad cuando se pierde o se desorganiza. Una experiencia característica del envejecimiento es la sensación de vértigo que a veces nos embarga cuando nos preguntamos de improviso en qué mes estamos. ¿Es el semestre de primavera o el semestre de otoño? Un caso más espectacular se da con la sensación de desconcierto que nos asalta al caminar en medio de la noche por un lugar desconocido. ¿Dónde diablos estoy?

9. *Conciencia activa y pasiva*

Quienquiera que reflexione sobre sus experiencias conscientes advertirá una distinción obvia entre la experiencia de la actividad intencional voluntaria, por un lado, y la experiencia de la percepción pasiva, por otro. No se trata, a mi juicio, de una distinción marcada, porque en la percepción hay un elemento voluntarista y la acción voluntaria tiene componentes pasivos. Pero sí existe una clara diferencia entre, por ejemplo, levantar los brazos voluntariamente como parte de un acto consciente y tenerlos alzados debido a que alguien ha estimulado nuestras conexiones nerviosas. La distinción está nítidamente ilustrada en las investigaciones del neurocirujano canadiense Wilder Penfield. Penfield

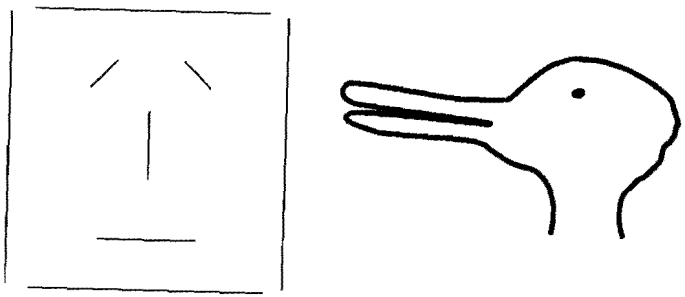
descubrió que mediante la estimulación de la corteza motriz de sus pacientes podía suscitar el movimiento de sus miembros. Indefectiblemente, el paciente decía: "Yo no lo hice, fue usted"[3]. En este caso, el paciente percibe el movimiento del brazo pero no hace la experiencia de la acción voluntaria. La distinción básica es la siguiente: en el caso de la percepción (ver el vaso frente a mí, sentir la camisa contra el cuello) tengo la sensación, percibo esto, y en ese sentido esto me pasa a mí. En el caso de la acción (levantar el brazo, caminar alrededor de la habitación) tengo la sensación, hago esto, y en ese sentido hago que esto suceda.

La convicción de la existencia de nuestro libre albedrío proviene, más que de cualquier otra cosa, de la experiencia de la acción voluntaria, que toda descripción de la mente debe tener en cuenta. Me explayaré más sobre el libre albedrío en el capítulo 8.

10. *La estructura* gestáltica

Nuestras experiencias conscientes no se nos presentan como un cúmulo desorganizado; antes bien, suelen hacerlo con estructuras bien definidas y a veces hasta precisas. En la visión normal, por ejemplo, no vemos manchas y fragmentos indiferenciados: vemos mesas, sillas, personas, autos, etc., aun cuando sólo fragmentos de esos objetos reflejen fotones en la retina y la imagen retinal esté distorsionada en diferentes aspectos. Los psicólogos *gestálticos* investigaron esas es-

3 W. Penfield, *The Mystery of the Mind: A Critical Study of Consciousness and the Human Brain*, Princeton, Princeton University Press, 1975, p. 76 [traducción española: *El misterio de la mente: estudio crítico de la conciencia y del cerebro humano*, Madrid, Pirámide, 1977].



tructuras y descubrieron algunos hechos interesantes. Uno de ellos es que el cerebro tiene la capacidad de tomar estímulos degradados y organizarlos en totalidades coherentes. Por lo demás, es capaz de recibir un estímulo constante y tratarlo en diferentes momentos como si fueran distintas percepciones. Así, en el famoso ejemplo del "pato-conejo" hay un aporte perceptivo constante, pero yo lo percibo unas veces como un pato y otras como un conejo.

En estos dibujos, aunque la figura de la izquierda no se parezca físicamente a una cara humana, la percibiremos como tal porque nuestro cerebro organiza el estímulo degradado en un todo coherente. La figura de la derecha es el célebre pato-conejo, que puede ser visto ora como el primero, ora como el segundo.

Por otra parte, la estructura *gestáltica* no sólo tiene que ver con la organización de nuestras percepciones en totalidades coherentes; dentro de todo el campo consciente, también hacemos una distinción entre las figuras que percibimos y el fondo sobre el cual se perciben. Así, por ejemplo, veo la pluma contra el fondo del

libro, el libro contra el fondo del escritorio, el escritorio contra el fondo del piso y el piso contra el fondo del resto de la habitación, hasta llegar al horizonte de todo mi campo perceptivo.

La estructura *gestáltica* de la conciencia, entonces, tiene al menos dos aspectos. En primer lugar, la capacidad del cerebro de organizar la percepción en totalidades coherentes, y segundo, su aptitud para diferenciar las figuras y el fondo.

11. El sentido del yo

Hay otra característica de las experiencias normales de la conciencia que no puedo dejar de mencionar. En ellas es típico que yo tenga cierta idea de quién soy y una sensación de mí mismo como un yo. Pero ¿qué podrá significar esto? No experimento mi "yo" de la misma manera que los zapatos en los pies o la cerveza que bebo. Soy incluso reacio a plantear esta cuestión, en primer lugar porque el debate sobre el yo tiene una sórdida historia en la filosofía, y segundo, peor aún, porque el problema del yo suscita interrogantes tan arduos que me cuesta abordarlos en este libro. Sin embargo, a la larga deberé enfrentarlos, de modo que reservo todo un capítulo, el 11, para una exposición sobre el yo.

Podríamos seguir enumerando rasgos, pero con los listados hasta aquí espero haber logrado transmitir la complejidad de nuestras experiencias conscientes. A continuación encontraremos motivos para destacar el rasgo esencial de la conciencia, a saber, la subjetividad cualitativa unificada, y nos será preciso explorar su relación con la intencionalidad.

II. Algunos otros enfoques filosóficos del problema de la conciencia

A lo largo del libro ya he analizado una serie de tratamientos de la filosofía de la mente, desde el materialismo eliminativo hasta el dualismo de las sustancias. De manera explícita o implícita, todos ellos son teorías de la conciencia. Por ejemplo, la teoría computacionalista de la mente se limita a decir que la conciencia es un proceso computacional en el cerebro. Es importante señalar que esa teoría, junto con otras formas de reduccionismo, no dice, por ejemplo, que si contáramos con el programa informático apropiado, la máquina, *por añadidura*, sería consciente. Dice, antes bien, que eso es todo lo que hay de conciencia. No hay otra cosa que el programa informático apropiado con sus entradas y productos pertinentes[4]. Sin embargo, a pesar de que he abarcado aquí muchas filosofías, aún quedan por mencionar una serie de concepciones influyentes de la conciencia. Por lo tanto, en interés de la exhaustividad, voy a examinar algunos puntos de vista que hasta ahora no he considerado.

1. Mistéricos

Los mistéricos estiman que la conciencia es un misterio imposible de resolver con nuestros métodos científicos actuales; algunos creen, además, que nunca podremos entender la explicación de la conciencia a través de los procesos cerebrales. Thomas Nagel[5] considera posible que alguna vez comprendamos cómo hace el cerebro para generar la conciencia, pero para ello será menester una revolución total en nuestro modo de pensar la realidad y nuestra concepción de la explicación científica, porque con el aparato actual no estamos en condiciones de entender cómo pueden los fenómenos neuronales de tercera persona dar origen a experiencias internas subjetivas y cualitativas. Colin McGinn[6], un mistérico radical, juzga imposible por principio que los seres humanos puedan comprender alguna vez el surgimiento de la conciencia a partir de la acción del cerebro.

Me parece que los mistéricos son demasiado pesimistas. Quizás acierten, por supuesto, al decir que nunca encontraremos una descripción científica de la conciencia. Pero la renuncia anticipada sería una muestra de derrotismo. Supongamos que encontramos efectivamente los diversos correlatos neuronales del campo consciente unificado. Y que, como segundo paso, podemos demostrar que esos elementos correlacionados son de hecho causas. Suponemos entonces, por así decirlo, que podemos encender la conciencia encendiendo esos procesos neurobiológicos, y apagarla si los apagamos. Como tercer paso debemos suponer que desarrollamos una teoría sobre el funcionamiento de todo el sistema. Esto es, presumimos nuestra capacidad de incorporar los enunciados de correlaciones causales a los enunciados de leyes o principios generales. Me parece que ese es precisamente el tipo de estructura teórica que hemos aceptado en otros ámbitos de la ciencia. La teoría de los

4 D. Dennett, *Consciousness Explained*, Boston, Little, Brown, 1991. Específicamente, el autor dice que la conciencia es una máquina virtual de Von Neumann implementada en una arquitectura conexionista.

5 T. Nagel, *The View from Nowhere*, Oxford, Oxford University Press, 1986 [traducción española: *Una visión de ningún lugar*, Madrid, Fondo de Cultura Económica, 1996].

6 C. McGinn, "Can We Solve the Mind-Body Problem?", *Mind*, 98, 1989, pp. 349-356.

gérmenes de la enfermedad es un buen ejemplo: en primer lugar, encontrar una correlación; segundo, comprobar que se trata en efecto de una correlación causal, y tercero, formular una teoría. Nagel objeta todo proyecto de esas características con el argumento de que aun cuando obtuviéramos esa correlación y pudiéramos proponer enunciados generales sobre ella, no alcanzaríamos el tipo de necesidad que cabe esperar de las explicaciones causales. Cuando explicamos, por ejemplo, por qué la mesa es sólida, podemos entender que, dado ese comportamiento molecular, la mesa *debe* resistir la presión de otros objetos y *debe* ser impenetrable por ellos. Ese "debe", cree Nagel, es típico de las explicaciones científicas.

En mi opinión, ese sentido de la necesidad es en gran medida una ilusión generada por las analogías que trazamos entre el comportamiento molecular y los objetos conocidos que nos rodean. Creemos que la mesa debe sostener los objetos porque consideramos que los movimientos moleculares forman una especie de rejilla del tipo con el cual estamos familiarizados. Pero las explicaciones de la ciencia no tienen como rasgo general la transmisión de cierta idea intuitiva de que las cosas deben ocurrir necesariamente así. Al contrario, la naturaleza es radicalmente contingente. Muchos de los principios explicativos más importantes de las ciencias distan de ser intuitivos u obvios. Piénsese en la ecuación de Schrödinger o la constante de Planck o, ya que estamos, la famosa fórmula de Einstein, $e = mc^2$. En cada caso, así resultó ser la naturaleza. No tenía por qué ser así, pero en los hechos resultó de ese modo. Coincido con Hume en pensar que la convicción de que la naturaleza debe ser *necesariamente* como es no es más que una ilusión. Así, por ejemplo, aun cuando una bola

de billar golpea otra, que la segunda se mueva es sólo un dato de la naturaleza. Pero también lo sería que ambas retrocedieran o que la primera tragara la segunda. Ocurrió, simplemente, que la naturaleza resultó de una manera y no de otra. La naturaleza está llena de sorpresas. Nunca debemos olvidar, por ejemplo, que el helio líquido 3 puesto en un recipiente trepa por las paredes de este. Por eso, la objeción de Nagel no me parece concluyente en absoluto con respecto a la posibilidad de una explicación neurobiológica de la conciencia.

2. *Superveniencia*

Decir que un fenómeno A superviene sobre un fenómeno B significa decir que A depende por completo de B de tal manera que cualquier cambio en la propiedad A debe correlacionarse con un cambio en la propiedad B. Por lo común se dice que la conciencia superviene sobre los procesos cerebrales. La idea básica es que no puede haber cambios en los estados mentales sin cambios correspondientes en los estados cerebrales. Por ejemplo, si paso de un estado en el que tengo sed a otro en que no la tengo, debe haber algún cambio correspondiente en mi cerebro. Y esto es verdad en general, de modo que los estados conscientes son totalmente dependientes de los estados cerebrales o supervienen sobre ellos. Varios filósofos han expuesto esta concepción; Jaegwon Kim fue tal vez quien lo hizo de manera más destacada[7]. La concepción lleva a una perspectiva

7 J. Kim, "Epiphenomenal and Supervenient Causation", en P. A. French, T. E. Uehling, Jr. y H. K. Wettstein (comp.), *Causation and*

a veces denominada "materialismo no reductivo". La idea de la superveniencia es proporcionar una descripción completamente materialista sin tratar de eliminar en ningún aspecto la conciencia. Esta doctrina se limita a decir que la conciencia superviene enteramente sobre los procesos cerebrales. Algunos han estimado que la superveniencia resuelve el problema mente-cuerpo o al menos muestra los primeros pasos en el camino a su solución.

Es sin duda cierto que la conciencia superviene sobre el cerebro. Pero este principio tiene una utilidad bastante limitada para la comprensión de las relaciones entre la mente y el cuerpo. Ello se debe a que hay dos tipos diferentes de superveniencia: la constitutiva y la causal. En filosofía, el concepto de superveniencia se utilizó tradicionalmente para describir las propiedades éticas y otras propiedades evaluativas. Se decía que dos actos no podían diferir exclusivamente en su bondad. No podía suceder que uno fuera bueno y el otro fuera malo y no existiera entre ellos otra diferencia. La bondad y la maldad debían supervenir sobre algunos otros rasgos del acto. Esto es lo que llamo "superveniencia constitutiva". Los rasgos que hacen a la bondad de un acto *no son la causa* de que este sea bueno; antes bien, *constituyen* su bondad. Pero esta analogía no se traslada a la mente de la manera como los filósofos partidarios de la superveniencia creyeron que lo haría. La superveniencia de la conciencia en los procesos cerebrales es de carácter causal. Esos procesos son causalmente responsables del rasgo que superviene. En el nivel de las activaciones neuronales, no constituyen la conciencia;

Causal Theories, Minneápolis, University of Minnesota Press, 1984, col. "Midwest Studies in Philosophy", vol. 9, pp. 257-270.

al contrario, las activaciones neuronales en el nivel inferior son la causa del rasgo sistémico o de nivel superior de la conciencia. Sin embargo, si esto es correcto, y todos nuestros conocimientos acerca del cerebro sugieren que lo es, el concepto de superveniencia no agrega nada a los conceptos ya existentes: la causación –incluida la causación de abajo arriba–, los niveles superior e inferior de descripción y los rasgos de orden superior que se realizan en el sistema compuesto de elementos del nivel inferior. La conciencia superviene sobre los procesos cerebrales, sí, pero ahora sigue siendo preciso decir cómo funciona.

3. *Panpsiquismo*

La doctrina del panpsiquismo sostiene que la conciencia está en todas partes. Esta concepción no suele enunciarse de manera explícita, pero está implícita en varios autores, sobre todo entre los místicos, quienes creen que si pretendemos explicar la conciencia en términos de microprocesos, alguna forma de ella ya debe estar presente de un modo u otro en estos. En un momento Thomas Nagel se dejó tentar por este punto de vista, y David Chalmers[8] lo explora y respalda, aunque no manifieste una adhesión explícita a él. Para esta perspectiva todo es consciente en alguna medida. Al dar un ejemplo de la ubicuidad de la conciencia, Chalmers describe con elocuencia en qué podría consistir ser un termostato consciente.

Al margen de su improbabilidad intrínseca, el panpsiquismo tiene el demérito adicional de ser inco-

8 D. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, *op. cit.*

herente. No veo de qué modo puede abordar el problema de la unidad de la conciencia. Esta no está diseminada como mermelada sobre un pedazo de pan, sino que aparece en unidades discretas. Si el termostato es consciente, ¿qué pasa con sus partes? ¿Hay una conciencia independiente para cada tornillo? ¿Para cada molécula? Si es así, ¿cómo se relaciona su conciencia con la conciencia de todo el termostato? Y si no es así, ¿qué principio hace que la unidad de la conciencia esté en el termostato y no en sus partes o en la totalidad del sistema de calefacción del cual aquel forma parte, o en el edificio donde está instalado ese sistema?

4. Neurobiología

Un cuarto conjunto de enfoques del tema que no he examinado hasta ahora está constituido por los intentos neurobiológicos de resolver el problema científico de la conciencia. A esta altura, no será un secreto para el lector que a mi juicio este enfoque es precisamente el apropiado. La investigación es tan importante que dedicaré a ella la siguiente sección.

III. Enfoques neurobiológicos actuales de la conciencia

Durante mucho tiempo, la mayoría de los neurobiólogos se mostraron renuentes a abordar el problema de la conciencia; en rigor, muchos aún son reacios a hacerlo. Las razones varían. Algunos sienten que "no están preparados" para estudiar la conciencia y que antes es necesario saber más sobre las funciones cerebrales en los fenómenos no conscientes. Otros creen que el problema de la conciencia no es realmente un

problema científico: debe dejarse en manos de teólogos y filósofos, pero en verdad no se lo conceptualiza como una cuestión científica. Un tercer grupo considera que no podemos plantear una descripción biológica de la conciencia, y que la ciencia nunca podrá explicar por qué la calidez se siente cálida o el rojo parece rojo. Adviértase la conexión entre este tipo de escepticismo y la concepción de los místericos que mencioné antes.

No obstante, nuestra época es notable por el gran número de neurobiólogos muy capacitados que intentan imaginar exactamente de qué modo los procesos cerebrales causan los estados conscientes. En un plano ideal, un proyecto de investigación con ese objetivo está compuesto por las tres etapas de las que hablé antes. Primero, encontrar el correlato neuronal de la conciencia, llamado CNC; segundo, verificar si la correlación es causal, y tercero, formular una teoría.

Para los fines de nuestro análisis, me parece que podemos dividir esa investigación en dos campos diferentes que denominó respectivamente "enfoque de los elementos constituyentes" y "enfoque del campo unificado". El primero considera que todo el campo consciente está compuesto de unidades conscientes más o menos independientes que yo denomino "elementos constituyentes". La experiencia del rojo, el sabor de la cerveza y el sonido del do mayor serían ejemplos del tipo de elementos constituyentes que tengo en mente. La idea de este enfoque es la siguiente: si pudiéramos representarnos con exactitud el modo como el cerebro causa aunque sólo sea un elemento constituyente, digamos la percepción del rojo, podríamos usar ese conocimiento para resolver todo el problema de la conciencia. Presuntamente, si podemos imaginar de qué manera el cerebro nos hace salvar la distancia entre el estímulo

entrante de la rosa roja y la experiencia visual consciente y real de la rojez, estaremos en condiciones de aplicar esas lecciones a otros colores, así como a sonidos, sabores, olores y a la conciencia en general. El enfoque de los elementos constituyentes parece idealmente apropiado para el proyecto de investigación de tres etapas que acabo de describir, y gran parte de las más interesantes investigaciones recientes representan un esfuerzo por encontrar el CNC de experiencias conscientes específicas.

A mi entender, es justo decir que la mayoría de los neurobiólogos dedicados al problema de la conciencia adhiere hoy a alguna versión del enfoque de los elementos constituyentes. Y sin duda es muy tentador suponer que deberíamos adoptar una perspectiva atomista sobre la conciencia, descomponer el problema en toda una serie de problemas mucho más pequeños y tratar de resolverlos uno a uno. No intentemos preguntarnos en general de qué manera el cerebro produce la conciencia; preguntémonos, en cambio, cómo produce la experiencia específica de la rojez de la rosa. Este enfoque atomista ha funcionado tan bien en el resto de la ciencia que parece natural suponer que sucederá lo mismo en el caso de la conciencia.

En la concepción de los elementos constituyentes suelen desplegarse tres líneas de investigación. En primer lugar, la investigación de la llamada vista ciega parece darnos una cuña ideal de entrada al problema de la conciencia. Los pacientes con vista ciega tienen daños en el área visual 1, situada en la parte posterior del cerebro. Pueden ver con normalidad en la mayor parte del campo visual, pero son ciegos en un segmento determinado. Sin embargo, a menudo son capaces de responder preguntas sobre sucesos que ocurren en ese

segmento del campo visual donde la ceguera los afecta. (De allí el uso de un aparente oxímoron: "vista ciega".) Así, por ejemplo, el paciente puede informar que hay una X o una O en la pantalla, aunque también diga que en realidad no la ve. Según dice, sólo "lo supone". Pero esas conjeturas tienden a ser acertadas una abrumadora mayoría de las veces, por lo cual no son una cuestión de azar. En cuyo caso podríamos, al parecer, encontrar el punto del cerebro en el cual la experiencia consciente de X difiere de la experiencia de la vista ciega: podríamos descubrir el CNC de esa experiencia visual.

Una segunda línea de investigación tiene que ver con la llamada rivalidad binocular y conmutación *gestáltica*. Si se presenta a uno de los ojos una serie de líneas horizontales y al otro una serie de líneas verticales, el sujeto no suele tener la experiencia visual de una cuadrícula, sino que ve alternativamente unas y otras. Ahora bien, como el estímulo perceptivo es constante y la experiencia difiere, al parecer deberíamos estar en condiciones de encontrar el punto del cerebro en el cual el mismo estímulo constante pasa de producir la experiencia de las líneas horizontales a producir la experiencia de las líneas verticales. En apariencia, esto nos daría el CNC de esas formas de conciencia.

Observaciones similares pueden hacerse con respecto a los fenómenos *gestálticos*. En el caso del pato-conejo, el estímulo constante en el papel produce ora la experiencia de un conejo, ora la experiencia de un pato. Si pudiéramos encontrar el punto del cerebro donde la experiencia pasa del pato al conejo y a la inversa, cabe conjeturar que tendríamos el CNC de estas experiencias.

Por último, una línea muy importante de investigación consiste simplemente en seguir las entradas de

estímulos perceptivos al cerebro y procurar localizar el punto en el cual causan experiencias visuales conscientes. Hay en la actualidad una enorme cantidad de investigaciones en curso sobre la visión, y muchos investigadores consideran estar frente a un camino prometedor para descubrir de qué manera el cerebro causa la conciencia[9].

El segundo abordaje del problema de la conciencia, el enfoque del campo unificado, comienza por considerar con detenimiento el rasgo de la unidad subjetiva cualitativa que antes mencioné. Para este enfoque, el paradigma de la conciencia, el blanco inicial de la investigación, no es la experiencia del color rojo, sino todo el campo consciente de la subjetividad cualitativa unificada. El interrogante fundamental no es cómo produce el cerebro tal o cual elemento constituyente específico en el campo consciente, sino cómo produce, en primer lugar, todo ese campo consciente. ¿Cuál es la diferencia entre el cerebro consciente y el cerebro inconsciente, y de qué manera esa diferencia explica causalmente la conciencia?

Pensémoslo de esta manera: imaginemos que nos despertamos en una habitación oscura. Tal vez estemos completamente despiertos y alertas aunque tengamos datos sensoriales mínimos. Imaginemos que no hay estímulos visuales ni sonidos. No vemos ni oímos nada. El único dato perceptivo es el peso del cuerpo contra la cama y el de las cobijas contra el cuerpo. Pero, y esto es lo importante, podemos llegar a estar totalmente conscientes y alertas en una situación de datos perceptivos mínimos. Ahora, en este punto, nuestro cerebro

9 C. Koch, *The Quest for Consciousness: A Neurobiological Approach*, Englewood (Colo.), Roberts and Co., 2004.

ha producido un campo consciente completo, y lo que debemos entender es cómo lo ha hecho y de qué manera el campo existe en él. Imaginemos a continuación que nos levantamos en esa habitación oscura, encendemos la luz y nos movemos de uno a otro lado. ¿Estamos creando conciencia? En cierto sentido sí, porque ahora tenemos estados conscientes que antes no teníamos. Sin embargo, me gusta pensar la situación de esta manera: no estamos creando una nueva conciencia, sino modificando el campo consciente preexistente. De acuerdo con el modelo del campo unificado, deberíamos considerar que los datos perceptivos no crean elementos constituyentes de la conciencia sino protuberancias y valles en el campo consciente, que debe existir con anterioridad a nuestras percepciones.

A mi juicio, el enfoque del campo unificado tiene más probabilidades de resolver el problema de la conciencia que el enfoque de los elementos constituyentes. ¿Por qué? Este último enfoque podría hacerlo, y es sin duda el punto de vista preferido por la mayoría de los investigadores que trabajan en ese ámbito. Sin embargo, tiene algunas características inquietantes que me llevan a estimar improbable su éxito. En el caso de un sujeto totalmente inconsciente en otros aspectos, este enfoque pronosticaría que, si pudiéramos encontrar siquiera el CNC de un solo elemento constituyente, por ejemplo el de la experiencia del rojo, el sujeto tendría de improviso una experiencia consciente de ese color y nada más. Tendría un relámpago consciente de rojez y luego volvería a caer de inmediato en la inconsciencia. Esto es lógicamente posible, por supuesto, pero no parece nada probable si tenemos en cuenta lo que sabemos del cerebro. Para expresarlo con mayor crudeza, una experiencia consciente del rojo sólo puede ocurrir

en un cerebro que ya es consciente. *Debemos considerar que la percepción no crea la conciencia, sino que modifica un campo consciente preexistente.* Reparemos ahora en los sueños. Como mucha gente, yo sueño en colores. Cuando veo el color rojo en un sueño, no tengo un dato perceptivo que crea un elemento constituyente del rojo. Antes bien, los mecanismos del cerebro que crean todo el campo unificado de la conciencia onírica generan mi experiencia del rojo como parte de ese campo.

Como dije antes, la mayoría de los investigadores adoptan el enfoque de los elementos constituyentes, y a mi entender lo hacen, al menos en parte, porque les permite formular un proyecto de investigación más sencillo. Parece muy difícil estudiar cantidades masivas de activaciones neuronales sincronizadas que puedan producir conciencia en grandes sectores del cerebro como el sistema talamocortical. Resulta mucho más fácil estudiar formas particulares de conciencia, por ejemplo las experiencias de los colores.

Por ahora, la cuestión es muy incierta. En los próximos años veremos más investigaciones sobre la conciencia. Apuesto al enfoque del campo unificado, pero estoy preparado para que me demuestren mi error.

IV. La conciencia, la memoria y el yo

Dije que en el estudio de la conciencia es útil observar los casos clínicos o patológicos, porque nos recuerdan características de los casos comunes y corrientes que quizá pasáramos por alto si no los contrastáramos con los ejemplos patológicos. Dos ejemplos que ya he mencionado son la desconexión callosa y la vista ciega. A continuación, un caso cerca de casa. El 4 de enero de 1999 yo estaba esquiando con rapidez sobre un terreno helado en la pista KT 22 de Squaw Valley, California. Desde mi punto de vista interno y subjetivo, recuerdo haber pensado que la luz era pobre y costaba ver las salientes. Lo siguiente que recuerdo es que estaba sentado en el elevador y me preguntaba qué día era. ¿Ya había pasado la Navidad? ¿Y el Año Nuevo? Miré a la mujer sentada frente a mí, que tenía un abono de tres días para el elevador, con vigencia desde el 4 hasta el 6 de enero. Supe que era el 4 de enero. (¿Por qué el 4 y no el 5 o el 6? Lo supe, y nada más.)

Las personas que vieron mi caída dicen que los esquíes se clavaron, pero yo salí lanzado y aterricé de cabeza. Me las arreglé para levantarme, encontrar las gafas protectoras y los anteojos en la nieve, volver a ponerme los esquíes y bajar con mucha cautela el resto de la montaña. Pero no respondía a las preguntas ni los intentos de entablar una conversación. Llegué al pie de la montaña y subí una vez más al elevador antes de "volver en mí".

Hay un lapso de 15 minutos de mi vida del cual no tengo absolutamente ningún recuerdo. Durante ese período me comporté como si tuviera plena conciencia, aunque no de manera completamente normal. El interés del caso deriva de la siguiente cuestión: ¿fui consciente durante ese cuarto de hora? El ejemplo se asemeja mucho a los casos de Penfield en que los pacientes, durante una convulsión epiléptica correspondiente al *petit mal*, siguieron realizando las actividades que los ocupaban, como manejar el auto de regreso a la casa o tocar el piano, aun cuando estaban inconscientes. Antes confiaba en la descripción de Penfield; ahora, después de haber hecho mi propia experiencia, no estoy tan seguro. En mi caso estoy convencido de que estuve consciente durante ese lapso, con la salvedad de que era

incapaz de registrar en la memoria mis experiencias conscientes. No tengo absolutamente ningún recuerdo, pero creo que me comporté como no me habría comportado de no haber estado consciente, si bien no me encontraba ciento por ciento normal. En este caso tenemos un nivel inferior de conciencia no registrado en la memoria. (De paso, los exámenes médicos revelaron que tenía una concusión y un hematoma subdural, de los que me recuperé por completo. Ahora esquío con casco.)

V. Conclusión

De todos los temas discutidos en el presente libro, este es el que me produce la mayor sensación de incomodidad. La conciencia es un fenómeno tan pasmoso y misterioso que uno siempre siente que el esfuerzo mismo de describirla con las palabras corrientes no sólo está en cierto modo destinado al fracaso, sino que el propio intento revela una falla del discernimiento. El carácter general de la relación de la conciencia con el cerebro, y por lo tanto la solución general al problema mente-cuerpo, no es difícil de enunciar: la conciencia es causada por procesos de micronivel con sede en el cerebro, y se realiza en este como un rasgo sistémico o de nivel superior. Pero esta caracterización omite abordar la complejidad de la estructura y la naturaleza precisa de los procesos cerebrales en cuestión. Sentimos la tentación de trivializar la conciencia considerándola un mero aspecto de nuestra vida; y desde luego, biológicamente hablando es sólo un aspecto, pero en lo que respecta a nuestras experiencias vitales concretas es la esencia misma de nuestra existencia significativa. Si Descartes no hubiera destruido ya el significado de la

frase, podríamos decir: "la esencia de la mente es la conciencia". Si trato de describir las variedades de su conciencia, usted comprobará que describo las variedades de su vida. Uno de los extraños rasgos de la vida intelectual reciente ha sido la idea de que la conciencia –en el sentido literal de estados y procesos subjetivos cualitativos– no era importante; de alguna manera, no contaba. Si esa idea parece tan descabellada es, entre otros motivos, porque la conciencia es la condición misma de la importancia de cualquier cosa. Sólo un ser consciente puede forjarse un concepto como el de la importancia.

CAPÍTULO 6

LA INTENCIONALIDAD

En la filosofía de la mente, el problema de la intencionalidad sólo es superado por el problema de la conciencia en materia de presunta y tal vez de imposible dificultad. A decir verdad, la cuestión de la intencionalidad se asemeja a una imagen especular del problema de la conciencia. Así como se supone que es extremadamente difícil desentrañar cuántos fragmentos de materia pueden ser conscientes dentro del cráneo o podrían crear conciencia a través de sus interacciones, también lo es imaginar cómo pueden "referirse" o remitir a algo del mundo más allá de sí mismos o generar esa referencia por medio de sus interacciones. Consideremos un ejemplo: en este momento pienso que el Sol está a ciento cincuenta millones de kilómetros de la Tierra. Mis pensamientos se refieren o remiten definitivamente al Sol. No aluden a la Luna, mi auto en el garaje, mi perro Gilbert o el vecino de al lado. Ahora bien, ¿qué elementos del pensamiento le permiten llegar a un lugar tan lejano como el Sol? ¿Envío rayos mentales hacia él, así como él emite rayos de luz que llegan a la Tierra? A menos que haya algún tipo de conexión entre el Sol y yo, cuesta imaginar cómo podrían mis pensamientos extenderse hasta el astro. Y lo que vale para el Sol vale para cualquier objeto que yo pueda representar en mis creencias, deseos y otros estados intencionales. Así, por ejemplo, si pienso que César cruzó el Rubicón, mi pensamiento se refiere a César, y su contenido es que este cruzó el Rubicón. Pero entonces, ¿qué elementos de la sustancia existente dentro de mi cráneo la llevan a re-

montarse en la historia a un individuo y un río determinados y atribuir al primero la acción específica de cruzar el segundo?

Además del problema de la posibilidad de una cosa semejante, hay un problema conexo: cómo puedo estar tan seguro de que sucede precisamente así. Cuando me refiero a Julio César, ¿cómo puedo estar tan relamidamente seguro de que mis pensamientos apuntan a César y no, por ejemplo, a Marco Antonio, Augusto o mi perro Gilbert? Si arrojo una piedra en la oscuridad, tal vez no tenga la menor idea de dónde golpea, pero cuando lanzo mi referencia a lo invisible, a menudo tengo completa certeza del lugar al que apunta.

Para empeorar las cosas, al parecer puedo en ocasiones pensar en objetos que ni siquiera existen. Cuando era niño creía que Papá Noel llegaba en Nochebuena. ¿Mi creencia se refería a él? Así parece ser, en efecto; sin embargo, ¿cómo puede ser eso posible, si Papá Noel ni siquiera existe?

Adviértase que sólo un filósofo se haría estas preguntas. La filosofía comienza con una sensación de misterio y se pregunta por cosas que cualquier persona en su sano juicio consideraría demasiado obvias para preocuparse por ellas.

Adviértase, también, que no podemos explicar la intencionalidad de la mente diciendo que es como la intencionalidad del lenguaje. En el caso de este último, el enunciado "César cruzó el Rubicón" se refiere a César y dice que cruzó ese río. No puedo decir que una representación mental deduce su capacidad intencional del lenguaje, porque el mismo problema, desde luego, se presenta en el caso de este. ¿Cómo es posible que una mera frase, sonidos que salen de mi boca o marcas que escribo en un papel, pueda referirse a, versar sobre

o describir objetos y situaciones de dos mil años atrás o ubicados a 15 mil kilómetros de distancia? La intencionalidad del lenguaje debe explicarse en términos de la intencionalidad de la mente y no a la inversa. Pues los sonidos y las marcas sólo se refieren a los objetos y sucesos que he mencionado en virtud de que la mente les ha impuesto una intencionalidad. El significado del lenguaje es intencionalidad derivada y debe deducirse de la intencionalidad original de la mente.

Con respecto a la intencionalidad, es preciso abordar tres problemas. Primero, cómo es ella posible; segundo, dado que los estados intencionales son posibles, cómo se determina su contenido, y tercero, cómo funciona todo el sistema de la intencionalidad. La mayor parte de la literatura filosófica se refiere a las dos primeras cuestiones, pero a mi juicio la más interesante es la tercera. En este capítulo voy a tratar en primer lugar el problema de la posibilidad de la intencionalidad, para lo cual utilizaré mi método habitual consistente en desmitificar el fenómeno a fin de bajarlo de las nubes. Luego me ocuparé del tercer tópico y describiré la estructura de la intencionalidad, e incluiré una sección sobre las diferencias entre intencionalidad, con c, e intensionalidad, con s. Por último, concluiré con la segunda cuestión, cómo se determinan los contenidos de los estados intencionales. Los lectores familiarizados con la ciencia cognitiva reconocerán que cuando hablamos de la intencionalidad aludimos a lo que en esa disciplina se conoce como "información". Yo prefiero hablar de "intencionalidad", porque "información" padece de una ambigüedad sistemática entre un sentido mental genuinamente independiente del observador (por ejemplo, al mirar ahora por la ventana obtengo información sobre el tiempo) y un sentido no mental rela-

tivo a él (por ejemplo, los anillos en el tocón de un árbol contienen información sobre la edad de este). Esa ambigüedad también puede afectar a "intencionalidad", pero es más fácil de evitar y la confusión resulta menos probable.

I. ¿Cómo es posible la intencionalidad?

Al parecer, este problema es tan arduo como el de la conciencia, de modo que sus presuntas soluciones se asemejan mucho a las propuestas para este último problema.

La solución dualista consiste en decir que, como hay dos reinos diferentes, el mental y el físico, el primero tiene su propia clase de facultades de las que el segundo carece. El reino físico es incapaz de referir; el reino mental, por su parte, es esencialmente capaz de pensar, y el pensamiento implica referencia. Espero que sea evidente que la solución dualista no soluciona nada. Para explicar el misterio de la intencionalidad apela al misterio de la mente en general.

A mi entender, la solución filosófica contemporánea más común al problema de la intencionalidad se encuentra en alguna forma de funcionalismo. La idea es que la intencionalidad debe analizarse íntegramente en términos de relaciones causales. Esas relaciones causales se dan entre el ambiente y el agente y entre diversos sucesos ocurridos dentro de este último. De acuerdo con esta concepción, la intencionalidad no tiene nada de misterioso. Es una mera forma de causación. El único rasgo especial es que existen relaciones intencionales entre las entrañas cerebrales del agente y el mundo externo. A esta altura, no necesito decir al lector que la versión más influyente del funcionalismo es el

funcionalismo computacional o inteligencia artificial fuerte.

Para terminar, hay una visión eliminativista de la intencionalidad: en realidad, los estados intencionales no existen. La creencia en cosas semejantes es el mero residuo de una psicología popular primitiva, que una ciencia madura del cerebro nos permitirá superar. Una variante de la perspectiva eliminativista es lo que podríamos llamar "interpretativismo". En este caso se supone que las atribuciones de intencionalidad siempre son formas de interpretación planteadas por algún observador exterior. Una versión extrema de esta concepción es la idea de Daniel Dennett de que a veces adoptamos la "postura intencional": no deberíamos suponer que la gente tiene literalmente creencias y deseos; sólo se trata de que nos resulta útil verlo de ese modo con el fin de predecir su comportamiento[1].

No dedicaré mucho tiempo a criticar estas distintas descripciones de la intencionalidad porque ya he criticado las ideas centrales generales de estos argumentos en capítulos anteriores. Lo que quiero hacer, como hice con el problema de la conciencia, es tratar toda la cuestión con los pies sobre la tierra. Si se pregunta cómo es posible que algo tan etéreo y abstracto como un proceso de pensamiento pueda llegar al Sol, la Luna, César y el Rubicón, el planteamiento parece conducir a un problema muy difícil. Pero si lo formulamos de una manera mucho más simple: ¿cómo puede un animal tener hambre o sed?, ¿cómo puede un animal ver o temer algo?, parece mucho más fácil de desentrañar.

1 D. Dennett, "The Intentional Stance", en *Brainstorms: Philosophical Essays on Mind and Psychology*, Montgomery (Vt.), Bradford Books, 1978.

Hablamos, como lo hicimos con respecto a la conciencia, de una serie determinada de capacidades biológicas de la mente. Y lo mejor es comenzar con las capacidades biológicas primitivas, por ejemplo el hambre, la sed, la pulsión sexual, la percepción y la acción intencional. En el capítulo anterior expuse algunos de los detalles neurobiológicos a través de los cuales los procesos cerebrales causan la sensación consciente de sed. Pero al explicar de qué manera esos procesos cerebrales causan la sensación de sed, ya explicamos cómo pueden causar formas de intencionalidad, porque la sed es un fenómeno intencional. Tener sed es tener deseos de beber. Cuando la angiotensina 2 se introduce en el hipotálamo y desencadena la actividad neuronal que a la postre resulta en la sensación de sed, lo que se está produciendo es, *eo ipso*, una sensación intencional. Las formas básicas de la conciencia y la intencionalidad son causadas por el comportamiento de las neuronas y se realizan en el sistema cerebral, compuesto justamente de estas. Lo que vale para la sed vale para el hambre, el miedo, la percepción, el deseo y todo lo demás.

Una vez que desmitificamos el problema de la intencionalidad al sacarlo del nivel espiritual abstracto y llevarlo al plano concreto de la biología animal real, me parece que no queda ningún misterio irresoluble acerca de la posibilidad de que los animales tengan estados intencionales. Si comenzamos con casos tan simples y obvios como el hambre y la sed, la intencionalidad no es difícil de explicar en absoluto. Desde luego, las creencias, los deseos y las formas sofisticadas de procesos de pensamiento son más complejos y están más alejados de la estimulación inmediata del cerebro debida al impacto del ambiente que las percepciones o la sensación de hambre y sed. Pero aun ellos son causados por procesos cerebrales y se realizan en el sistema cerebral.

Cuando la mera existencia de las relaciones intencionales nos parece misteriosa y nos planteamos preguntas como la siguiente: ¿cómo es posible que mis pensamientos lleguen a puntos tan lejanos como el Sol o se remonten en la historia hasta épocas tan remotas como la de Julio César?, lo hacemos porque atribuimos un modelo erróneo de relaciones a las frases que describen nuestros contenidos intencionales. De manera similar, cuando nos desconcierta que podamos tener pensamientos sobre cosas que no existen en absoluto, como Papá Noel, nuestra perplejidad se debe a que concebimos la intencionalidad como si fuera una relación análoga al hecho de estar junto a usted, alcanzarlo o sentarme encima. Uno no puede alcanzar algo que no existe ni sentarse encima de un objeto que está a ciento cincuenta millones de kilómetros de distancia. Pero el hecho de referirse a algo o pensarlo no tiene nada que ver con sentarse sobre él o alcanzarlo. Se trata, antes bien, de una forma de *representación*, y el concepto de esta no exige que la cosa representada exista efectivamente o se encuentre en alguna proximidad inmediata a su representación. Deberíamos escuchar la pregunta: ¿cómo es posible pensar en Papá Noel si este ni siquiera existe?, como si nos preguntáramos: ¿cómo es posible inventar una historia sobre Papá Noel, si este ni siquiera existe? En este último caso el problema es más sencillo, pues advertimos que desde un punto de vista metafísico la invención de historias ficticias no es algo difícil. Cuando digo esto no resuelvo el problema, por supuesto, porque, estrictamente hablando, la intencionalidad de la historia deriva de la intencionalidad del contenido mental. Trato de disipar una sensación de

misterio mostrando que lo aparentemente misterioso es como lo obviamente nada misterioso. Nuestra aptitud de tener contenidos intencionales acerca de lo inexistente parece enigmática, pero la capacidad de construir relatos ficticios lo parece mucho menos.

Sin embargo, hay muchos otros problemas. Por ejemplo, ¿cuál es la relación entre la intencionalidad consciente e inconsciente, y cómo obtienen su contenido los estados intencionales? Tendré que abrirme camino hasta el punto en que pueda responder estas preguntas. Por ahora, me parece que lo mejor es describir la estructura formal de los estados intencionales, porque no captaremos el funcionamiento de la intencionalidad mientras no estudiemos los rasgos estructurales de esos estados, como las creencias y los deseos, las esperanzas y los temores, las percepciones, los recuerdos y las intenciones.

II. La estructura de la intencionalidad

1. Contenido proposicional y modo psicológico

Como los estados intencionales son capaces de referirse a objetos y estados de cosas en el mundo más allá de sí mismos, deben tener alguna clase de *contenido* que determine esa referencia; en efecto, es preciso distinguir el contenido del estado del tipo de estado de que se trata. Así, puedo creer que lloverá o esperar, temer o desear que llueva. El contenido es el mismo en los cuatro casos: que va a llover, pero se relaciona con el mundo de modos psicológicos diferentes: creencia, temor, esperanza, deseo, etc. Esta distinción, dicho sea de paso, es un paralelo exacto de la misma distinción en el lenguaje. Así como puedo ordenarte que salgas de

la habitación, puedo predecir que saldrás de ella y preguntar si vas a salir. Tenemos el mismo contenido en todos los casos: que vas a salir de la habitación, pero presentado en tipos diferentes de actos de habla. Una buena manera de pensarlo es considerar que el estado consiste en un modo psicológico, como la creencia o el deseo, con un contenido proposicional, como la proposición de que está lloviendo. Podemos representarlo como E(p), donde E simboliza el modo o tipo de estado y p el contenido proposicional. A menudo, esos estados se denominan "actitudes proposicionales".

No todos los estados intencionales tienen como contenido una proposición entera. Uno podría simplemente admirar a Eisenhower o amar a Marilyn, y en esos casos el estado intencional sólo se refiere a un objeto. Tales estados pueden representarse como E(n), donde n nombra un objeto o se refiere a él.

Adviértase que las representaciones intencionales siempre se muestran con ciertos aspectos y no otros. Por ejemplo, yo podría representar intencionalmente un objeto como el lucero del alba y no como el lucero de la tarde, aunque un único objeto sea ambas cosas. El aspecto "cuerpo celeste que brilla cerca del horizonte en el atardecer" no es el mismo que "cuerpo celeste que brilla cerca del horizonte a la mañana". *Los estados intencionales siempre tienen formas aspectuales*, por lo tanto toda representación aparece con determinados aspectos. Este es un detalle importante, pues toda teoría de la intencionalidad debe explicar la forma aspectual y algunas teorías materialistas son incapaces de hacerlo. En el capítulo 3 señalé que el funcionalismo no podía distinguir entre el deseo de agua y el deseo de H_2O, debido a que las relaciones funcionales en las cuales se apoya esa doctrina para analizar la intencionali-

dad no tienen las formas aspectuales de la auténtica intencionalidad. En el capítulo 9 veremos que cualquier teoría de lo inconsciente debe explicar la presencia de la forma aspectual cuando un estado intencional es inconsciente.

2. *Dirección del ajuste*

Los estados intencionales –como los actos de habla, otra vez– se relacionan con el mundo de diferentes maneras. La meta de una creencia es ser verdadera, y si lo es logra su cometido. Si es falsa, fracasa. Los deseos, por su parte, no presumen representar el mundo tal como es, sino como nos gustaría que fuera. Así, si creo que llueve, mi creencia será verdadera si y sólo si está lloviendo. Pero si deseo que llueva, satisfaré o cumpliré mi deseo si y sólo si llueve. Aunque ambas situaciones parezcan similares, hay una distinción crucial. En el caso de la creencia, se supone que el estado intencional representa el modo como las cosas son en el mundo. La creencia, por decirlo así, es *responsable de ajustarse al mundo*. En el caso del deseo, en cambio, su meta no es representar las cosas como son sino como querríamos que fueran. Aquí, por decirlo de alguna manera, *el mundo es responsable de ajustarse al contenido del deseo*. Voy a introducir algo de jerga para describir esta distinción. Cuando el estado mental es responsable de ajustarse a una realidad con existencia independiente, podemos decir que la *dirección* de su ajuste es *"de la mente al mundo"* o, de manera alternativa, que tiene una *responsabilidad* de ajuste *"de la mente al mundo"*. El estado mental se ajusta o no al modo como las cosas son realmente en el mundo. Las creencias, las convicciones, las hipótesis, etc., así como las experiencias

perceptivas, tienen esta dirección de ajuste de la mente al mundo. Las expresiones más comunes para evaluar el logro de esa dirección de ajuste son "verdadera" y "falsa". De las creencias y convicciones puede decirse que son verdaderas o falsas. Los deseos y las intenciones no lo son del mismo modo que las creencias, porque su meta no es concordar con una realidad de existencia independiente, sino conseguir que esta coincida con el contenido del estado intencional. Por esa razón diré que tienen una *dirección* de ajuste o son *responsables* del ajuste *"del mundo a la mente"*.

Algunos estados intencionales, aunque tienen un contenido proposicional, carecen de una dirección de ajuste porque su meta no es concordar con la realidad (la dirección de la mente al mundo) ni hacer que esta coincida con ellos (la dirección del mundo a la mente). Antes bien, dan por sentado que el ajuste ya existe. Así, si lamento haberte pisado el pie o me alegra que brille el sol, doy por sentado que te pisé el pie y que el sol brilla. En lo concerniente a esos casos, digo que los estados intencionales tienen una "dirección *nula* de ajuste". "Presuponen" una relación de ajuste en vez de afirmarla o tratar de provocarla. Me parece conveniente representar la dirección de ajuste de la mente al mundo con una flecha hacia abajo, de este modo: ↓; el ajuste del mundo a la mente con una flecha hacia arriba, ↑, y el ajuste nulo con el signo correspondiente: Ø.

3. *Condiciones de satisfacción*

Cada vez que tenemos un estado intencional con una dirección de ajuste no nula, el ajuste se alcanzará o no: la creencia será verdadera, el deseo se cumplirá, la intención se llevará a cabo o no, según corresponda.

En tales casos, podemos decir que la creencia, el deseo o la intención han sido satisfechos. En ese aspecto, la verdad de la creencia, el cumplimiento del deseo y la ejecución de la intención se corresponden. Propongo describir este fenómeno diciendo que todo estado intencional con una dirección no nula de ajuste tiene *condiciones de satisfacción*. Podemos concebir los estados mentales como representaciones de sus condiciones de satisfacción. En rigor, más adelante sostendré que estas son la clave para entender la intencionalidad, pero a fin de decirlo necesitamos algunos elementos más en nuestro aparato.

4. *Autorreferencialidad causal*

Los fenómenos intencionales más básicos desde el punto de vista biológico, incluyendo las experiencias perceptivas, las intenciones de hacer algo y los recuerdos, tienen un rasgo lógico peculiar en sus condiciones de satisfacción. Como parte de las condiciones de satisfacción de mi recuerdo de que ayer fui a un día de campo, por ejemplo, se cuenta el hecho de que, si realmente me acuerdo del suceso, este mismo debe causar mi recuerdo de él. Si detallamos las condiciones de satisfacción del recuerdo, estas no sólo son que el suceso haya ocurrido, sino también que su ocurrencia cause el recuerdo mismo que incluye esa ocurrencia en el resto de sus condiciones de satisfacción. Podemos describir esta situación diciendo que los recuerdos, las intenciones y las experiencias perceptivas son causalmente autorreferenciales. Lo cual significa que el contenido mismo del estado se refiere a este al hacer un requerimiento causal. Las condiciones de satisfacción del recuerdo exigen que la causa de este sea el suceso recordado. Las

condiciones de satisfacción de la intención requieren que la ejecución de la acción representada en el contenido de aquella exija que esa misma intención cause dicha ejecución. Y así sucesivamente en otros casos.

En este aspecto, las intenciones, los recuerdos y las experiencias perceptivas difieren de las creencias y los deseos. Podemos presentar la diferencia de la siguiente manera. Si creo que ayer fui a un día de campo, la estructura formal de mi estado intencional es esta:

Creer (que ayer fui a un día de campo).

Pero si recuerdo que ayer fui a un día de campo, la estructura formal de mi estado intencional es la siguiente:

Recordar (que ayer fui a un día de campo, y el hecho de ir a un día de campo causó ese recuerdo).

En los estados con una dirección de ajuste de la mente al mundo es preciso distinguir los que son causalmente autorreferenciales, como las percepciones y los recuerdos, de los que no lo son, como las creencias. En un paralelo exacto con ello, en los estados cuya dirección de ajuste es del mundo a la mente debemos diferenciar los que son autorreferenciales en términos causales, como la intención que tengo antes de hacer algo (lo que llamo "intención previa") y mi intención mientras lo hago efectivamente (lo que llamo "intención en la acción"), de los que no lo son, como los deseos. Además, todo estado causalmente autorreferencial con una dirección de ajuste también tiene una dirección de causación. En la percepción visual, por ejemplo, si veo que el gato está sobre el felpudo, sólo veo las cosas

como realmente son (y de ese modo logro una dirección de ajuste de la mente al mundo) si el hecho de que el gato esté sobre el felpudo me lleva a ver la situación de esa manera (dirección de causación del mundo a la mente). En la acción intencional la flecha apunta hacia el otro lado. Logro alcanzar intencionalmente el libro que está en el anaquel más alto (y obtengo así una dirección de ajuste del mundo a la mente) sólo si mi intento, mi intención en la acción, es la causa de mi éxito (dirección de causación de la mente al mundo).

Las relaciones formales resultantes son tan hermosas que no puedo resistir la tentación de presentarlas en un cuadro, donde utilizo la anticuada terminología de la cognición y la volición para denominar las dos familias:

	COGNICIÓN			VOLICIÓN		
	Percepción	Memoria	Creencia	Intención en la acción	Intención previa	Deseo
Auto-rreferencia causal	sí	sí	no	sí	sí	no
Dirección de ajuste	↓	↓	↓	↑	↑	↑
Dirección de causalidad	↑	↑	Ninguna	↓	↓	Ninguna

5. La red de intencionalidad y el contexto de las capacidades preintencionales

En general, los estados intencionales no se presentan en unidades aisladas. Si creo, por ejemplo, que está lloviendo, no puedo tener esa mera creencia aislada. Debo creer, por ejemplo, que la lluvia consiste en gotas

de agua, que estas caen del cielo, que por lo común bajan y no suben, que mojan el suelo, que provienen de nubes formadas en el cielo y así sucesivamente, de manera más o menos indefinida. Desde luego, alguien podría creer que está lloviendo y carecer de algunas de esas otras creencias, pero en general parece que la creencia de que llueve sólo es lo que es debido a su posición en una "red" de creencias y otros estados intencionales. Y podemos concebir que la totalidad de nuestros estados intencionales forma una elaborada red interactuante. Podemos decir incluso que un estado intencional sólo funciona –esto es, sólo determina sus condiciones de satisfacción– en relación con las redes de las cuales forma parte. Si creo ser dueño de un automóvil, también debo creer que los autos son medios de transporte, que se los utiliza en calles y carreteras, que van de un lado a otro, que las personas pueden subir y bajar de ellos, que los automóviles son un tipo de bien susceptible de comprarse y venderse, etcétera.

Si seguimos los hilos de la red, al final llegaremos a una serie de aptitudes, maneras de afrontar el mundo, disposiciones y capacidades en general que llamo colectivamente el "transfondo" [*Background*]. Por ejemplo, si abrigo la intención de ir a esquiar, sólo puedo hacerlo si doy por sentado que tengo la aptitud de esquiar, pero esta no es en sí misma una intención, creencia o deseo adicional. Sostengo la tesis polémica de que, para funcionar, los estados intencionales en general exigen un trasfondo de capacidades no intencionales.

He presentado un esbozo muy breve de la estructura formal de la intencionalidad. Podemos resumirla de la siguiente manera. En lo concerniente a cualquier estado intencional hay una distinción entre su tipo y su contenido. Cuando el contenido es toda una proposi-

ción, representará situaciones del mundo y lo hará con una de las tres direcciones de ajuste: de la mente al mundo, del mundo a la mente o nula. Así, los estados intencionales que no tienen una dirección nula de ajuste son representaciones de sus condiciones de satisfacción. Y dada la red de intencionalidad, aun los que tienen una dirección nula y los que carecen de un contenido proposicional completo están, con todo, constituidos en gran medida por estados que tienen efectivamente una dirección no nula de ajuste. De tal modo, si me disculpo por haberte pisado el pie, debo creer que lo hice y desear no haberlo hecho. Si admiro a Jimmy Carter debo tener un conjunto de creencias y deseos relacionados con él. En general, *la intencionalidad es representación de condiciones de satisfacción*. Los estados intencionales más básicos en términos biológicos, los que establecen una relación directa de los animales con su medio ambiente, tienen un componente causalmente autorreferencial en sus condiciones de satisfacción. Un estado intencional sólo puede funcionar, esto es, puede determinar condiciones de satisfacción, en virtud de su posición en una red de estados intencionales y dado el trasfondo de capacidades preintencionales.

Más adelante, al hablar del inconsciente en el capítulo 9, veremos que la red de intencionalidad, cuando es inconsciente, es en realidad un caso especial de aptitudes contextuales, la aptitud de producir fenómenos intencionales conscientes.

La estructura formal de la intencionalidad que he descrito no es un asunto trivial. Se trata, de hecho, de la estructura de nuestra vida consciente. En rigor, es la estructura de nuestra vida mental, tanto consciente como inconsciente. Cuando llegamos a entender una situación social en la cual nos encontramos inmersos,

cuando decidimos embarcarnos en algún curso de acción, cuando percibimos el cielo en una noche estrellada, cuando recordamos de improviso episodios de nuestra infancia mientras comemos una magdalena, estamos frente a manifestaciones de la estructura formal que he descrito. A fin de entender nuestra vida, debemos entender la estructura de la intencionalidad.

Es importante destacar que esta discusión no tiene pretensión alguna de hacer fenomenología. Hablamos de la estructura lógica de la intencionalidad. La fenomenología, en su mayor parte, es incapaz de acceder a esa estructura.

III. La intencionalidad, con c, y la intensionalidad, con s

El lector sólo entenderá la literatura filosófica actual sobre la intencionalidad si capta la diferencia entre la intencionalidad con c y la intensionalidad con s.

Aun los filósofos profesionales suelen confundirlas. La intencionalidad con c, como hemos visto, es la propiedad de la mente por la cual esta se dirige, se refiere o alude a objetos y situaciones del mundo independientes de sí misma. La intensionalidad con s es lo contrario de la *extensionalidad*. Se trata de una propiedad de ciertas frases, enunciados y otras entidades lingüísticas por la cual estas incumplen ciertas pruebas de extensionalidad. La conexión entre ambas radica en que muchas frases sobre estados intencionales, con c, son frases intensionales, con s. Hay varias pruebas de la extensionalidad, pero las dos más célebres son la de sustitución (a veces llamada ley de Leibniz) y la de inferencia existencial. Consideremos una y otra en orden. La prueba de sustitución dice que cada vez que dos ex-

presiones se refieren a lo mismo, podemos sustituir una por otra sin cambiar el valor de verdad del enunciado en el cual hacemos la sustitución. Formalmente podemos expresarlo de la siguiente manera:

1. $[(a = b) \ \& \ Fa] \rightarrow Fb$.
   Si a es idéntica a b y tiene una propiedad F, entonces b tiene la propiedad F.
   Así, de
2. César cruzó el Rubicón,
   y
3. César es idéntico al mejor amigo de Marco Antonio, podemos inferir
4. El mejor amigo de Marco Antonio cruzó el Rubicón.
   Por este motivo, se dice que la presencia de "César" en 2 es *extensional* con respecto a la sustituibilidad. Pero hay frases en las cuales no podemos hacer la sustitución. Así, de
5. Bruto cree que César cruzó el Rubicón,
   y la proposición de identidad 3, no podemos inferir válidamente
6. Bruto cree que el mejor amigo de Marco Antonio cruzó el Rubicón,
   porque Bruto quizá no crea que César es el mejor amigo de aquel. Se dice que esta frase es *intensional* con respecto a la presencia de César. No pasa la prueba de sustituibilidad.
   El principio de existencia inferencial dice que siempre que a tiene la propiedad F, puede inferirse válidamente la existencia de algún objeto con esa misma propiedad F.
7. $Fa \rightarrow (\exists)(Fx)$
   Así, de

8. John vive en Kansas City,
   podemos inferir válidamente
9. Hay algún x tal que John vive en x.
   Pero hay frases de esta forma en las que no podemos dar por válida la inferencia. Así, de
10. John busca la ciudad pérdida de la Atlántida,
    no se sigue que
11. Hay algún x tal que John busca x.
    Porque la ciudad que busca tal vez ni siquiera exista.

Las frases del tipo de la número 10 se caracterizan como intensionales, porque no pasan la prueba de la inferencia existencial.

Nótese que las dos frases intensionales se refieren a estados intencionales con c. Esto ha llevado a algunos filósofos a suponer erróneamente que la intencionalidad tiene por esencia algo intensional. Pero están equivocados. La razón por la cual las frases sobre estados intencionales con c son a menudo intensionales con s es la siguiente: los estados mismos son representaciones de sus condiciones de satisfacción. Pero las frases acerca de dichos estados no son representaciones de esas condiciones, sino representaciones de sus representaciones. De allí que su verdad o falsedad no dependa de cómo son las cosas en el mundo real según las representan los estados intencionales originales, sino de cómo son en el mundo de las representaciones tal como este existe en la mente de los agentes cuyos estados intencionales se representan. Así, cuando digo que César cruzó el Rubicón, hablo sin duda de César y el Rubicón. Pero cuando digo que Bruto cree que César cruzó el Rubicón, hablo de Bruto y de lo que le sucede en la cabeza. La verdad de mi dicho no depende del mundo real de

César y el Rubicón sino de lo que en la cabeza de Bruto representa a uno y a otro. No puedo hacer entonces la sustitución a menos que tenga una premisa adicional con el propósito de que Bruto la acepte. Observaciones análogas son válidas para la prueba de la inferencia existencial. Si hablo del lugar donde John vive efectivamente, hablo de una persona y un lugar reales, pero si hablo de lo que John busca, me refiero a un estado intencional, el intento de encontrar algo, cuyas condiciones de satisfacción él trata de hacer realidad. Pero John podría tener ese estado intencional y buscar algo, aun cuando ese algo que busca no existiera. Una vez más, el hecho de que la frase intensional con s sea una representación de una representación explica su intensionalidad.

En lo concerniente a la distinción entre intencionalidad con c e intensionalidad con s, lo importante es recordar que la intencionalidad no tiene nada intrínsecamente intensional. Un enunciado en el sentido de que Bruto cree que César cruzó el Rubicón es en verdad un enunciado intensional con s. Pero no por ello lo es la creencia misma, la creencia real de Bruto. Esa creencia es tan extensional como puede serlo. Sólo será verdad si tanto César como el Rubicón existen (inferencia existencial) y algo idéntico al primero cruzó algo idéntico al segundo (sustituibilidad).

No pretendo dar la idea de que sobre la base de los párrafos anteriores el lector entenderá todo lo que puede entenderse acerca de la intensionalidad con s. Hay mucho más para decir. En mi libro *Intentionality: An Essay in the Philosophy of Mind* se encontrarán más detalles[2]. Todo lo que quiero hacer ahora es proporcionar al lector herramientas suficientes para seguir los argumentos sobre la intensionalidad con s y la intencionalidad con c sin cometer los errores que son corrientes en la filosofía contemporánea.

2 J. R. Searle, *Intentionality: An Essay in the Philosophy of the Mind*, Cambridge, Cambridge University Press, 1983 [traducción española: *Intencionalidad: un ensayo en filosofía de la mente*, Madrid, Tecnos, 1992].

IV. La determinación del contenido intencional: dos argumentos en apoyo del externalismo

La mayoría de los filósofos dedicados a estos temas parecen creer que hay una pregunta muy general, con una respuesta igualmente general, de la forma: ¿cómo se determina el contenido de nuestros estados intencionales? Se supone que la pregunta no debe interpretarse como: ¿cuál es la explicación de que lleguemos a tener estos contenidos intencionales y no otros?, sino de la siguiente manera: ¿cómo se *constituyen* los contenidos intencionales? ¿Qué elemento del estado intencional tal como este existe aquí y ahora hace que sea un deseo de agua y no de otra cosa? Por curioso que parezca, aunque se trata de preguntas muy diferentes, la concepción más influyente en la actualidad considera que una respuesta a la primera –¿cuál es la explicación causal de que tengamos estos estados intencionales?– proporciona una respuesta a la segunda: ¿qué hecho de estos estados intencionales los constituye con el contenido que tienen? Esta concepción, denominada "externalismo", dice que el contenido intencional está constituido en gran medida por las relaciones causales

(externas) del agente con el mundo externo, y no por los rasgos (internos) de la mente o el cerebro.

El punto de vista que he adoptado tácitamente a lo largo de este libro es una forma de internalismo. De acuerdo con el internalismo así concebido, nuestros contenidos intencionales están íntegramente vinculados a lo que tenemos dentro de la cabeza. Se refieren, por supuesto, a objetos y situaciones del mundo. Para eso está la intencionalidad: para relacionarnos con el mundo mediante la representación de sus diversos rasgos. El contenido que permite a un estado intencional referirse a un objeto y no a otro se encuentra en su totalidad entre uno y otro oídos del sujeto referente. Así concebido, el internalismo ha sido cuestionado en décadas recientes por una serie de argumentos favorables a la idea de que los contenidos mentales no están en la cabeza, o al menos no lo están del todo, sino que residen en gran medida en las relaciones entre lo que sucede en ella y el resto del mundo. Es importante advertir que esta teoría externalista no se limita a afirmar que nuestros contenidos mentales internos suelen ser causados por sucesos externos (ambas partes coinciden en ello); sostiene, antes bien, que esos mismos contenidos no son verdaderamente internos sino, a lo sumo, una mezcla de interioridad y exterioridad. Si el lector encuentra vaga esta postura, me temo que acierta, porque el externalismo es una tesis planteada con bastante vaguedad. A continuación esbozaré los dos principales argumentos sostenidos en defensa del externalismo, lo cual nos ayudará a disipar un poco sus oscuridades. Para explicar esos argumentos debo introducir la noción de indexicalidad. Una frase o expresión indexical se refiere a algún objeto indicando las relaciones que este mantiene con el enunciado mismo de la expresión.

De modo que si yo digo "tengo hambre" y usted dice "tengo hambre", enunciamos la misma frase con el mismo significado, pero los enunciados tienen diferentes condiciones de satisfacción debido a la aparición del indexical "yo". El "yo" enunciado por mí se refiere a mí. El "yo" enunciado por usted se refiere a usted. En el lenguaje hay muchas formas de indexicalidad: "yo", "tú", "aquí", "ahora", "esto", "aquello", "ayer", "mañana" y "por allí", así como los tiempos de los verbos, son ejemplos de indexicales.

El primer argumento en apoyo del externalismo: Hilary Putnam y la Tierra Gemela[3]

Tal vez cabría considerar que el "agua" puede definirse como un líquido transparente, incoloro e insípido presente en lagos y ríos y procedente del cielo en forma de lluvia. Sin embargo, dice Hilary Putnam, eso no nos da el significado de "agua". Para verlo, imaginemos una galaxia como la nuestra, con un planeta idéntico al nuestro, que llamaremos Tierra Gemela. En la Tierra Gemela todo es exactamente igual que en la Tierra, molécula por molécula, con una sola excepción. Lo que en la Tierra llamamos "agua" está compuesto de H_2O; lo que los habitantes de la Tierra Gemela llaman "agua" no es H_2O sino una fórmula química muy larga que podemos abreviar como "XYZ". Ahora bien, en

3 H. Putnam, "The Meaning of 'Meaning'", en K. Gunderson (comp.), *Language, Mind, and Knowledge*, Minneápolis, University of Minnesota Press, 1975, pp. 131-193, fragmento reeditado en D. Chalmers (comp.), *The Philosophy of Mind*, *op. cit.* [traducción española: "El significado del 'significado'", en Luis Valdés Villanueva (comp.), *La búsqueda del significado: lecturas de filosofía del lenguaje*, Madrid, Tecnos, 1995].

1750, antes de que nadie supiera nada sobre la composición química, lo que había en la cabeza de los habitantes de la Tierra Gemela cuando utilizaban la palabra "agua" era exactamente igual a lo que había en la cabeza de los terrícolas cuando usaban la misma palabra. De todas maneras, si bien los contenidos de las cabezas eran iguales, los significados eran diferentes. Los significados no pueden estar en la cabeza, porque en sus cabezas hay las mismas cosas que en las nuestras, pero los significados difieren. En la Tierra, "agua" se refiere a un tipo de sustancia; "agua", en la Tierra Gemela, se refiere a otro tipo de sustancia. Tanto en uno como en otro planeta, dice Putnam, el significado es determinado por relaciones causales que los hablantes mantienen con sustancias presentadas de manera indexical. En la Tierra, "agua" significa todo lo que tenga la misma estructura que esta sustancia indexicalmente presentada. Otro tanto para la Tierra Gemela. Pero como las sustancias son diferentes, H_2O en un caso, XYZ en otro, los significados también lo son. Los significados, concluye Putnam, "sencillamente no están en la cabeza"[4].

Lo que vale para el significado vale para el contenido mental en general. Las creencias que utilizan la palabra "agua" no son iguales para la gente de la Tierra Gemela y los habitantes de la Tierra. Pero de ser así, debemos deducir que las creencias no pueden estar íntegramente en la cabeza. En esta hay exactamente lo mismo en ambos casos, pero las creencias son diferentes.

4 H. Putnam, "The Meaning of 'Meaning'", en D. Chalmers (comp.), *The Philosophy of Mind*, *op. cit.*, p. 587.

El segundo argumento en apoyo del externalismo: Tyler Burge y la artritis[5]

Tyler Burge ha presentado un argumento conexo para mostrar que, al menos en parte, los contenidos de la mente son sociales. Así reza su planteamiento. Imaginemos que Joe va a ver a su médico en Santa Mónica, y dice: "Doctor, me duele el muslo. Creo que es artritis". Podemos suponer que el médico responde: "Si el dolor es en el muslo, no puede ser artritis. La artritis es una inflamación de las articulaciones". Imaginemos ahora que el estado de Joe es exactamente el mismo, pero la comunidad es diferente. En la cabeza de nuestro hombre hay exactamente lo mismo, porque se trata de la misma persona en el mismo momento. Digamos, empero, que no está en Santa Mónica sino en Santa Mónica Gemela. E imaginemos que en esta comunidad la palabra "artritis" se utiliza de otra manera: designa los dolores musculares y las inflamaciones articulares. Ahora bien, en este segundo caso el contenido del cerebro de Joe es exactamente el mismo que en el primero, pero su creencia, al parecer, es diferente. En Santa Mónica cree erróneamente tener artritis. En Santa Mónica Gemela su creencia es verdadera. No podemos presentar esta creencia diciendo que Joe cree tener artritis, porque *arthritis* ["artritis"] es una palabra del inglés normal. En Santa Mónica no hablan este idioma, al menos en lo concerniente a esa palabra. Por lo tanto, debemos inventar otra. Podemos decir que en San-

5 T. Burge, "Individualism and the Mental", en P. A. French, T. E. Uehling, Jr. y H. K. Wettstein (comps.), *Studies in Metaphysics*, Minneápolis, University of Minnesota Press, 1979, col. "Midwest Studies in Philosophy", vol. 4; extracto reeditado en D. Chalmers (comp.), *The Philosophy of Mind*, *op. cit.*

ta Mónica nuestro hombre sostiene una creencia verdadera, la de que tiene *tartritis*. Ahora bien, y este es el sentido del experimento mental, aunque en los dos casos el contenido de su cabeza es exactamente el mismo (y debe serlo porque Joe es exactamente la misma persona en el mismo momento), hay no obstante dos creencias diferentes. Deben ser dos creencias diferentes porque una es verdadera y la otra falsa, y una misma creencia no puede ser ambas cosas a la vez.

La conclusión es similar a la de Putnam. Así como este mostraba que los significados están constituidos en parte por relaciones causales con el mundo, el argumento de Burge demuestra que los contenidos mentales están parcialmente constituidos por relaciones sociales con la propia comunidad. En ambos casos hemos demostrado, al parecer, que los contenidos intencionales no son internos a la cabeza.

¿Qué debemos hacer con estos argumentos? Admiro la perspicacia filosófica de sus autores, pero me parece que los dos planteamientos son falaces. La idea básica del internalismo es que la mente –y por "mente" aludimos aquí a lo que está dentro de la cabeza– fija condiciones que un objeto debe cumplir a fin de que una expresión u otra forma de contenido mental pueda hacer referencia a él. En un ejemplo clásico, la expresión "el lucero del alba" fija una condición tal que, si un objeto la satisface, la expresión puede utilizarse literalmente para referirse al objeto. En la exposición de Putnam no hay ningún elemento que cuestione esta concepción. Este filósofo sustituye la idea tradicional de que una lista de rasgos se asocia a cada palabra –por ejemplo, a la palabra "agua" se asocian características como transparente, incolora, líquida, etc.– por una definición indexical: "El agua es cualquier cosa de es-

tructura idéntica a lo que vemos ahora". Según nuestra descripción de la autorreferencialidad causal de la intencionalidad perceptiva, eso equivale a decir que el agua es cualquier cosa de estructura idéntica a la sustancia que causa esa misma experiencia visual. Pero esa definición establece una condición que está enteramente representada en los contenidos de la mente. Los terrícolas ven una sustancia que llaman "agua" y fijan una condición que será cumplida por cualquier cosa que tenga una similitud importante con el elemento que han bautizado con aquel nombre. En cuanto a los habitantes de la Tierra Gemela, podemos contar exactamente la misma historia. Ven una sustancia que denominan "agua" y establecen una condición que será satisfecha por cualquier cosa con una similitud relevante. La condición es completamente interna a los contenidos de la mente. El hecho de que una sustancia la satisfaga o no depende del mundo y no de la mente, exactamente del mismo modo que cualquier otra condición fijada en el plano interno, como ser el lucero del alba, cuyo cumplimiento o incumplimiento por parte de un objeto también dependerá del mundo y no de la mente. El internalismo es una teoría que nos dice de qué manera la mente fija condiciones. La referencia a los objetos corresponde cuando estos las satisfacen. Las condiciones que se establezcan dependen de la mente; que un objeto las satisfaga depende del mundo. No he visto nada en las críticas externalistas que ponga en tela de juicio esta idea básica.

En el caso del ejemplo de Burge, la única diferencia de los estados mentales de Joe en una y otra circunstancia es de carácter indexical. En ambas comunidades él cree lo siguiente:

1. Tengo este dolor en el muslo. Creo que es artritis.
   Pero también tiene un supuesto previo contextual que podemos expresar así:
2. Doy por sentado que mi uso de las palabras coincide con el de la comunidad, y cuando haya una diferencia modificaré mi uso para coincidir con ella.
   Sin embargo, una aplicación de 2 al presente caso resulta en:
3. Doy por sentado que en mi comunidad "artritis" se refiere a dolores como este; si no es así, modificaré mi uso para adecuarme a la comunidad.

De tal modo, en cualquier uso de un lenguaje público interviene un componente indexical. La diferencia entre Joe en el primer caso y Joe en el segundo es que la comunidad es diferente. En el primer caso, nuestro hombre se equivoca con respecto a 3. Ese tipo de dolores no se denominan "artritis". En mi opinión, este ejemplo no plantea problema alguno, ni siquiera para las versiones más ingenuas del internalismo. En respuesta a esta objeción, Burge me ha dicho (en una conversación) que pretende sencillamente estipular que Joe no tiene creencias metalingüísticas sobre el modo de utilizar las palabras. De acuerdo. No hace falta suponer que Joe ha pensado la cuestión en absoluto. Pero uno de los supuestos contextuales de nuestro uso de las palabras es que compartimos significados con otros miembros de nuestra comunidad. Cuando Joe descubre que ese supuesto contextual es erróneo, no modifica de ninguna manera su concepción de los hechos no lingüísticos –aún tiene el mismo dolor en el mismo lugar–, pero sí su uso lingüístico. Burge tiene razón, me parece, al considerar posible y razonable suponer que nuestro

hombre nunca pensó de manera explícita que su uso se ajusta a la comunidad. Pero el supuesto previo sobre el carácter comunitario del uso lingüístico es un supuesto contextual general, algo anterior a las creencias y pensamientos explícitos. Se presume que nuestro uso del lenguaje se adecua a los demás miembros de nuestra comunidad; si no fuera así, no podríamos tener la pretensión de comunicarnos con ellos a través de un lenguaje compartido.

V. El contenido mental interno y su manera de relacionar a los agentes con el mundo

A fin de explicar con mayor profundidad los errores de estas objeciones al internalismo, debo hablar un poco de la naturaleza del contenido mental y su modo de relacionar a los agentes con el mundo. Ya hemos visto que un estado intencional fija condiciones de satisfacción. Así, por ejemplo, si creo que Sócrates toma agua, mi creencia será cierta, y por lo tanto quedará satisfecha, si y sólo si Sócrates bebe agua. Las preguntas que nos hacemos ahora son: ¿qué rasgos constituyen los componentes del pensamiento de que Sócrates bebe agua, y cómo relacionan esos elementos componentes al agente con el pensamiento total y el mundo externo? En este caso, centremos la atención en "Sócrates" y "agua". (Dejaré al margen la discusión del "bebe", porque la predicación plantea problemas especiales que van más allá de las cuestiones del externalismo y el internalismo.) Todo el mundo coincide en que cada uno de los componentes, "Sócrates" y "agua", hace un aporte a la condición total de verdad del pensamiento. "Sócrates" alude a Sócrates y "agua" se refiere al agua. Así como la condición de verdad de que Sócrates toma agua

está asociada a toda la frase, cada uno de estos dos componentes tiene una condición asociada, la de que uno y otro contribuyan a la condición de verdad de la frase en su totalidad. Hay entonces dos conjuntos de cuestiones sobre los componentes del pensamiento. En primer lugar, cómo se relaciona cada elemento con la condición que él determina, y segundo, cómo se relaciona el agente con la determinación de esas condiciones. Si admitimos que "Sócrates" se refiere a Sócrates y "agua" se refiere al agua, ¿cómo debe el agente relacionarse con estas palabras a fin de poder usarlas para determinar las condiciones de satisfacción de todo el pensamiento? La respuesta tradicional, y la proporcionada por el sentido común, es que cada palabra fija las condiciones que fija debido a su *significado*, y el agente puede usarlas como las usa porque *conoce* el significado de cada una de ellas. Y el conocimiento del significado le permite utilizar la palabra de tal manera que puede incluir la condición correspondiente en las condiciones de verdad de toda la frase.

Podemos enunciar ahora la disputa entre los internalistas y los externalistas con un poco más de precisión: ambas partes coinciden en que las palabras hacen un aporte a las condiciones de verdad de toda la frase y en que hay cierta condición que el propio hablante debe satisfacer a fin de poder utilizar esas palabras para fijar las condiciones de verdad en cuestión. La disputa se refiere por entero a la naturaleza de la condición cumplida por el hablante. El interrogante es el siguiente: ¿la condición asociada a la palabra es algo que se representa en la mente o el cerebro del hablante o algo parcialmente independiente de estos? Según el internalista, la condición debe estar representada en la cabeza del hablante. A criterio del externalista, los contenidos

de la cabeza son insuficientes para hacer una referencia cabal. A eso aludía Putnam cuando decía: "los significados sencillamente no están en la cabeza". El argumento propuesto por los externalistas es el mismo en todos los casos: dos hablantes podrían tener en la cabeza contenidos idénticos en su tipo, pero significar algo diferente. En cambio, la respuesta dada a esta idea por los internalistas es: siempre que sucede así, se debe a que en la cabeza hay algún componente indexical que fija una condición diferente de satisfacción en uno y otro caso, porque la establece con referencia a la cabeza del hablante en cuestión. Por ejemplo, si suponemos que dos gemelos idénticos que lo son, según suele decirse, "molécula por molécula", piensan "tengo hambre", cabe estimar que los contenidos mentales son idénticos en su tipo, pero de todos modos quieren decir algo diferente porque el gemelo A se refiere a sí mismo y el gemelo B se refiere a sí mismo. La indexicalidad permitirá que pensamientos de tipo idéntico en la cabeza determinen diferentes condiciones de satisfacción, porque estas, al determinarse indexicalmente, se fijan en relación con la cabeza en cuestión. Así, en el caso de la Tierra Gemela los habitantes de esta y de la Tierra fijan condiciones de satisfacción relativas a sí mismos: lo que llamamos "agua" es algo cuya estructura es de tipo idéntico a la sustancia que *nosotros* vemos. Pero como en ambos casos el "nosotros" es diferente y las personas de la Tierra Gemela ven algo diferente de los terráqueos, tendrán diferentes condiciones de satisfacción aun cuando los contenidos de la cabeza sean idénticos en su tipo. En este ejemplo nada muestra que los significados no están en la cabeza.

Observaciones análogas pueden plantearse con respecto al ejemplo de Burge. Joe tiene exactamente el

mismo pensamiento en las dos comunidades. Ese pensamiento es: "Tengo este dolor. Creo que es artritis". Y el supuesto previo contextual es que los dolores como este se llaman "artritis" en mi comunidad. Pero como la comunidad es diferente en uno y otro caso, el mismo pensamiento determinará diferentes condiciones de satisfacción en relación con las dos comunidades. En un caso Joe tiene una creencia verdadera; en otro, tiene una creencia falsa.

Volvamos a la cuestión original. Si rechazamos la tesis externalista de que el contenido intencional es determinado por cadenas causales externas, ¿qué es entonces lo que lo determina? Si hablamos en términos causales, no creo que haya ninguna respuesta general a esta cuestión, salvo decir que nuestros contenidos intencionales están determinados por una combinación de nuestras experiencias vitales y nuestras capacidades biológicas congénitas. Ya he esbozado una explicación de la determinación de la sensación de sed del animal por procesos neurobiológicos. Si cambiáramos ligeramente el ejemplo, de manera que yo no tuviese sed en general sino de un vaso de cerveza de malta irlandesa de barril o de una copa de Chateau Lafitte de 1953, la historia sería mucho más complicada. Tendría que explicar por qué mis experiencias vitales me llevaron a hacer cierta clase de experiencias relacionadas con el sabor, que fui capaz de evocar en la memoria, así como pude forjar el deseo de repetirlas en el futuro. Pero si la historia tiene que ser más complicada para explicar un deseo específico, llegaría a serlo de manera increíble si yo tratara de describir cómo podría haberme formado una intención cuyo contenido fuese escribir la gran novela norteamericana, casarme con una republicana o exponer la intencionalidad en un solo capítulo.

Sin embargo, si no hablamos de la historia de nuestros estados intencionales sino de su *constitución*, por ejemplo, qué hechos en mí me llevan a la creencia de que César cruzó el Rubicón, deberemos apelar a la noción de condiciones de satisfacción.

Antes de abordar directamente la cuestión, recapitulemos para ver dónde estamos. Comenzamos el capítulo con tres preguntas:

1. ¿Cómo es posible la intencionalidad?
2. ¿Cómo se determinan los contenidos intencionales?
3. ¿Cuál es el funcionamiento en detalle de los estados intencionales?

No hicimos tanto contestar la primera pregunta como suprimir la necesidad de plantearla en ese tono de voz filosófico especial que hace imposible cualquier respuesta. La bajamos de los cielos transformándola en interrogantes como este: ¿cómo es posible para un animal tener sed, hambre o miedo? Una vez contestados estos interrogantes queda respondida la primera pregunta, en cuanto es una pregunta con significado. Dejamos a un lado la segunda pregunta hasta responder la tercera. De pasada, rechacé la respuesta externalista a esa segunda pregunta. Ahora quiero utilizar los resultados obtenidos al contestar la tercera para realizar en la segunda el mismo tipo de maniobra hecha en la primera. La pregunta: ¿cómo me es posible tener una creencia cuyo contenido es que César cruzó el Rubicón?, no es en principio más difícil de responder que esta otra: ¿cómo me fue posible tener sed de agua, esto es, tener un deseo cuyo contenido es que beba agua? En ambos casos la respuesta radica en ver la conexión

esencial entre intencionalidad y condiciones de satisfacción. Lo que hace de mi deseo un deseo de tomar agua es que lo satisfaré si y sólo si tomo agua. Este no es un pronóstico psicológico sobre lo que me hará sentir bien, sino la definición del contenido intencional relevante. Exactamente de la misma manera, lo que hace que mi creencia tenga el contenido "César cruzó el Rubicón" es el hecho de que se satisfará si y sólo si César cruzó el Rubicón. El contenido del estado intencional es precisamente lo que lleva a este a tener las condiciones de satisfacción que tiene. Esas condiciones de satisfacción siempre se representan bajo ciertos aspectos. Yo represento a un hombre determinado como César, por ejemplo, y no como el mejor amigo de Marco Antonio, aun cuando César sea idéntico al mejor amigo de Marco Antonio.

Sin embargo, ¿no es circular esta respuesta a la segunda pregunta? ¿Qué hace que un estado intencional tenga el contenido que tiene? Respuesta: el hecho de tener las condiciones de satisfacción que tiene. ¿Y cuáles son esas condiciones de satisfacción? Las determinadas por el contenido del estado intencional. Esto parece circular, sin duda. Pero se trata precisamente de la clase de circularidad que busco. No aceptamos la cuestión tal como está planteada; antes bien, la rechazamos para sustituirla por una descripción del funcionamiento real de la intencionalidad. Esta funciona en virtud de la existencia de conexiones muy rigurosas entre contenido intencional, forma aspectual y condiciones de satisfacción. El paso siguiente para anclar toda esta descripción en el mundo real consiste en señalar el papel central de la conciencia. Tener conscientemente un estado intencional, por ejemplo pensar de manera consciente que César cruzó el Rubicón, es estar conscientemente al tanto de las condiciones de satisfacción. Tener inconscientemente el mismo estado intencional es tener algo que al menos en principio es susceptible de volverse consciente. En el capítulo 9 analizaré de manera pormenorizada la relación entre lo consciente y lo inconsciente. Por ahora, me basta con decir lo siguiente. Rechazamos la tercera pregunta formulada en el sentido que no admite ninguna respuesta y la reemplazamos por una explicación del funcionamiento real del contenido intencional. Este funciona efectivamente porque los agentes intencionales tienen pensamientos conscientes cuya identidad misma es tal que puede determinar la vigencia de determinadas condiciones de satisfacción y no de otras. Esas condiciones de satisfacción se representan según ciertos aspectos y no otros. Si preguntamos: ¿cómo puede un estado de mi cerebro tener el contenido "César cruzó el Rubicón"?, la cuestión parece imposible de resolver. En cambio, si preguntamos: ¿cómo puede mi pensamiento consciente "César cruzó el Rubicón" tener el contenido de que César cruzó el Rubicón?, ya no parece imposible responderla. Conozco los significados de las palabras, sé cómo se relacionan con objetos y situaciones del mundo y al formarme todo el pensamiento soy consciente de que tiene precisamente esta condición de satisfacción: César cruzó el Rubicón. Una vez que rechazamos el sentido metafísico de la tercera pregunta, la asimilamos a una descripción general del modo real de funcionamiento de la intencionalidad y de esa manera la desmitificamos. Y eso es todo lo que hace falta decir acerca de la constitución del contenido intencional en general. Más allá de eso, desde luego, es preciso decir mucho –y en gran parte ya lo he dicho– sobre la red y el trasfondo, la dirección del ajuste

y la autorreferencialidad causal, el modo psicológico y todo lo demás.

Expondré las relaciones entre conciencia e intencionalidad en el capítulo 9. Por el momento, sólo esto: una enorme ventaja evolutiva de la conciencia humana radica en nuestra capacidad de coordinar una gran cantidad de intencionalidad ("información") de manera simultánea en un sólo campo consciente unificado. Piénsese en la cantidad de intencionalidad coordinada ("procesamiento de información") existente cuando, por ejemplo, manejamos el auto a la mañana para ir al trabajo. No se tome en cuenta exclusivamente la coordinación de la percepción y la acción. (Por ejemplo, paso al automóvil de mi derecha. Adelante hay una luz roja.) Considérese también el acceso constante de intencionalidad inconsciente, por ejemplo: llegaré tarde a mi cita de las nueve de la mañana; ¿dónde voy a almorzar?; me pregunto cómo saldrán las reuniones. Se trata de representaciones intencionalistas del mundo, y por su conducto afrontamos este último.

VI. Conclusión

Dije al comienzo de este libro que lo peor que podemos hacer es dar al lector la impresión de que entiende algo que en realidad no entiende. No quiero que con la lectura de este capítulo crea haber comprendido la intencionalidad. Apenas he raspado la superficie de un tema muy amplio. Sí deseo, en cambio, que el lector tenga cierta concepción global de la intencionalidad como representación y pueda evitar errores que son comunes en la filosofía contemporánea. Específicamente, es preciso ver la distinción entre intencionalidad con c e intensionalidad con s. Deben advertirse las dificultades existentes en las descripciones externalistas hoy ortodoxas del contenido intencional, y es necesario comenzar a captar la conexión entre intencionalidad y conciencia, que explicaré en detalle en el capítulo 9. Y sobre todo, el lector debe empezar a hacerse una idea del funcionamiento de la intencionalidad como un rasgo real del mundo real, comprensión que le permitirá, espero, evitar sentirse intimidado y pensar que en la intencionalidad intrínseca u original hay algún profundo misterio inaccesible a toda explicación natural.

CAPÍTULO 7

LA CAUSALIDAD MENTAL

Uno de los problemas residuales heredados del dualismo es el de la causación mental. Nuestro primer problema mente-cuerpo era: ¿cómo pueden los procesos físicos causar de algún modo procesos mentales? Pero para muchos filósofos la otra mitad de la cuestión es aún más acuciante: ¿cómo puede algo tan etéreo e insustancial como los procesos mentales causar de alguna manera efectos físicos en el mundo real? Con seguridad, el mundo físico real está "causalmente cerrado", en el sentido de que nada exterior a él puede tener efectos causales en su interior.

A esta altura, el lector sabrá que, a mi entender, no se trata de interrogantes de imposible resolución; lo que los hace parecer arduos es nuestra aceptación de las categorías cartesianas. Sin embargo, en el estudio de la causación mental surgen muchos problemas fascinantes. Aun cuando el lector acepte mi descripción general de las relaciones entre la mente y el cuerpo, creo que en el análisis presentado en este capítulo encontrará algunas cuestiones interesantes sobre esa causación.

I. Hume y su explicación de la causación

Debemos comenzar con Hume. Así como cuando hablamos de la mente en general no hay manera de escapar a Descartes, cuando hablamos de la causación no podemos eludir a Hume. Su explicación de la causación es, con mucho, su aporte filosófico más original, vigoroso y profundo, y creo que la mayoría de los filó-

sofos coincidirían conmigo en que se trata de uno de los textos de filosofía más impresionantes jamás escritos en lengua inglesa. Cualesquiera sean las demás enseñanzas que el lector extraiga de este libro, me gustaría que aprendiera algo sobre la escéptica exposición de la causación presentada por Hume. (Lo que sigue no pretende ser, desde luego, un sustituto de la lectura del original, la tercera parte del primer libro del *Tratado* de Hume; no obstante, lo que diré a continuación puede servir como guía para explorar ese territorio[1].) Allá vamos:

Hume comienza por preguntarse cuáles son los componentes de nuestro razonamiento al considerar la causa y el efecto. En el siglo XXI expresaríamos la cuestión de esta forma: ¿cuál es la definición de "causa"? Nuestro concepto de causación, dice Hume, tiene tres componentes:

1. Prioridad, esto es, la necesidad de que la causa ocurra con anterioridad; las causas no pueden venir después de sus efectos.
2. Contigüidad en el espacio y el tiempo, con lo cual se refiere a que la causa y el efecto deben ser adyacentes. Si me rasco la cabeza en Berkeley y un edificio se derrumba en París, el hecho de haberme rascado no puede ser la causa del derrumbe, a menos que haya una serie de eslabones en una "cadena causal" entre mi cabeza y el edificio parisino.
3. Conexión necesaria: además de la prioridad y la contigüidad, la causa y el efecto deben estar conectados por necesidad, de tal manera que la primera *produzca* realmente el segundo, lo *haga suceder*, lo *necesite* o, como resume Hume, que haya una *conexión necesaria* entre causa y efecto.

1 D. Hume, *A Treatise on Human Nature*, edición establecida por L. A. Selby-Bigge, Oxford, Clarendon Press, 1951 [traducción española: *Tratado de la naturaleza humana*, Barcelona, Orbis, 1981].

Sin embargo, dice Hume, cuando empezamos a estudiar casos reales, comprobamos que no podemos encontrar ninguna conexión necesaria. Observamos, por ejemplo, que cuando toco el interruptor la luz se enciende, y cuando vuelvo a tocarlo, se apaga. Creo que hay una conexión causal entre el toque del interruptor A y la luz que se apaga en B, pero en realidad lo único que puedo observar es A seguido de B. Hume presenta la ausencia de conexión necesaria como si se tratara de una especie de lamentable falta que podríamos superar si hiciéramos una inspección más detenida. Pero sabe perfectamente bien que, del modo como ha descrito el caso, esa conexión nunca podría existir. Supongamos, en efecto, que yo dijera que la conexión necesaria entre el toque del interruptor y el encendido de la luz es el pasaje de electricidad a través del cable C, y descubriera algún método de observarlo, digamos a través de un dispositivo de medición. Pero eso no serviría. Pues ahora tendría el toque del interruptor, el pasaje de la electricidad y el encendido de la luz, la secuencia ACB, pero ninguna conexión necesaria entre esos tres sucesos. Y si encontrara alguna, si descubriera aparentes conexiones necesarias entre el interruptor A, la electricidad C y la luz B, con la forma, por ejemplo, del cierre del circuito D o la activación de las moléculas en el filamento de tungsteno E, no se trataría, de todos modos, de conexiones necesarias. Tendría entonces una secuencia de cinco sucesos, ADCEB, que exigirían conexio-

nes necesarias entre sí. La primera conclusión escéptica de Hume es que no existe conexión necesaria entre la llamada causa y el llamado efecto.

En este punto, nuestro filósofo realmente emprende el vuelo. Dice que debemos examinar los principios subyacentes de la causa y el efecto, y descubre dos: el principio de causación y el principio de causalidad. El primero afirma que todo suceso tiene una causa. El segundo dice que a iguales causas, iguales efectos. Hume ve atinadamente que no se trata de principios equivalentes. Pues podría ocurrir que todo suceso tuviera una causa y no hubiera coherencia en el tipo de efectos de una causa específica ni en el tipo de causas de un efecto determinado. Podría ser, asimismo, que cuando hubiera causas y efectos, iguales causas tuvieran iguales efectos, aunque no todos los sucesos tuviesen una causa. Pero, dice Hume, si examinamos estos dos principios, el principio de causación y el principio de causalidad, encontramos un rasgo singular. No parecen ser demostrables. No son verdaderos por definición. Es decir, no son verdades analíticas. Deben ser, entonces, verdades empíricas sintéticas. Pero en ese caso, y esto es lo decisivo del argumento de Hume, no hay manera de establecerlos mediante métodos empíricos, porque cualquier intento de establecer algo a través de esos métodos presupone justamente esos dos principios.

Esta es la conclusión más célebre de Hume. Recibe el nombre de problema de la inducción y a continuación veremos cómo se formula. Pensemos en argumentos deductivos, como el siguiente:

Sócrates es hombre.
Todos los hombres son mortales.
Por lo tanto, Sócrates es mortal.

Puede advertirse que el argumento es válido porque la conclusión ya está contenida de manera implícita en las premisas. En aquella no hay nada que no esté en estas. Podríamos representarlo mediante un diagrama y decir que vamos de la premisa a la conclusión, P → C, donde P ≥ C. La premisa siempre contiene más información que la conclusión (o en un caso restrictivo en el cual deducimos una proposición de sí misma, la conclusión es igual a la premisa). La validez está garantizada porque en la conclusión no hay nada que ya no esté en las premisas. Pero cuando consideramos los argumentos científicos o inductivos, como el elaborado para probar nuestra premisa de que todos los hombres son mortales, no tenemos al parecer este tipo de validez. Pues en el caso de estos argumentos vamos de la evidencia E a la hipótesis H. Decimos, por ejemplo, que la evidencia sobre la mortalidad de determinados hombres proporciona evidencia para, o respalda, o establece la hipótesis general de que todos los hombres son mortales. Pasamos de la evidencia a la hipótesis, E → H, pero (y aquí está la diferencia con respecto a la deducción) en el caso de la inducción siempre hay más en la segunda que en la primera. La hipótesis siempre es algo más que un mero resumen de la evidencia. Es decir, E < H, E es menos que H. En tal caso, podría parecer vergonzoso utilizar siquiera una vez los argumentos inductivos, pero estos son, desde luego, absolutamente esenciales; ¿de qué otro modo, en efecto, estableceríamos las proposiciones generales que forman las premisas de nuestros argumentos deductivos? ¿Cómo podríamos acaso establecer que todos los hombres son mortales si no pudiéramos generalizar a partir de instancias específicas de hombres mortales, o de otros ti-

pos de evidencia sobre casos particulares, para llegar a la conclusión general de la mortalidad de todos?

Cuando pasamos de la evidencia a la hipótesis, cuando decimos que la primera respalda la segunda, la establece o la confirma, no lo hacemos de una manera arbitraria o injustificada. Al contrario, tenemos algunos principios o reglas R en virtud de las cuales pasamos de una a otra, y podríamos considerarlas como las reglas del método científico. Entonces, no establecemos arbitrariamente E → H, sino que pasamos de E a H sobre la base de R: ER → H. Ahora bien, y aquí tenemos el planteo crucial de Hume, ¿cuál es el fundamento de R? Supondremos que E, la evidencia, proviene de observaciones reales, y H es una generalización de estas. Pero en tal caso, si debemos justificar el paso de E a H sobre la base de R, ¿cuál es la justificación de R? Hume responde: cualquier intento de justificar R presupone R. ¿Qué es R exactamente? (En este punto aparece la conexión con la causación y la causalidad.) R puede formularse de diversas maneras. La más obvia es decir simplemente que todo suceso tiene una causa y causas iguales tienen iguales efectos. También puede decirse que los casos no observados se asemejarán a los casos observados, que la naturaleza es uniforme o que el futuro se parecerá al pasado. Hume estima todas esas aserciones como más o menos equivalentes para estos fines. Si no presuponemos algún tipo de uniformidad de la naturaleza, la uniformidad garantizada por la causalidad y la causación, no tenemos fundamentos para plantear argumentos inductivos. Pero, y esto es lo crucial, la creencia en la uniformidad de la naturaleza no tiene fundamento, porque cualquier creencia semejante debería fundarse en la inducción, que a su turno tendría que fundarse en la uniformidad de la naturaleza; así, el

intento de basar la creencia en dicha uniformidad sería circular.

Hasta aquí, las conclusiones de Hume son casi totalmente escépticas. En la naturaleza no hay conexiones necesarias, y tampoco existe una base racional para la inducción. En una actitud característica de su método, luego de llegar a conclusiones escépticas Hume nos da razones por las cuales no podemos aceptarlas y debemos proceder como si el escepticismo no se hubiese establecido. Estamos condenados a continuar con nuestras viejas supersticiones, y Hume está ávido de explicarnos exactamente de qué manera.

Cuando buscamos conexiones necesarias no encontramos ninguna que se sumara a la prioridad y la contigüidad, pero sí dimos con otra relación: la conjunción constante de instancias semejantes. Descubrimos que la cosa que llamamos causa siempre es seguida por la cosa que llamamos efecto. Como un mero dato de nuestra existencia en el mundo, descubrimos que las cosas que denominamos causas siempre son seguidas por las cosas que denominamos efectos. Esta repetición constante en nuestra experiencia, esa conjunción permanente de instancias semejantes, da origen a cierta expectativa en nuestra mente, en virtud de la cual cuando percibimos la cosa que llamamos causa, automáticamente esperamos percibir la cosa que llamamos efecto. Esta "determinación sentida de la mente" de pasar de la percepción de las causas a las expectativas vívidas del efecto, y de la idea de la causa a la idea del efecto, suscita en nosotros la ilusión de que en la naturaleza hay algo más que prioridad, contigüidad y conjunción constante. Esa determinación sentida de la mente nos da la convicción de que en la naturaleza hay conexiones necesarias. Dicha convicción, sin embargo, no es más

que una ilusión. La única realidad es la realidad de la prioridad, la contigüidad y la conjunción constante. Según la explicación de Hume, la causación sólo es, literalmente, una condenada cosa tras otra. Con la única salvedad de que hay una regularidad en el modo como una cosa sigue a otra, y esa regularidad nos da la ilusión de que existe algo más. Pero la conexión necesaria que a nuestro juicio hay en la naturaleza es una completa ilusión de la mente. La única realidad es la regularidad.

La existencia de la regularidad en casos previamente observados no es razón alguna, empero, para suponer que el caso siguiente se parecerá a los precedentes. No representa de ninguna manera una solución al problema de la inducción. Nos da la ilusión de poder resolver ese problema, porque creemos que con la determinación sentida de la mente hemos descubierto una conexión necesaria. Pero esa conexión está íntegramente en nuestra cabeza y no en la naturaleza misma. En sustancia, entonces, Hume afronta el problema de la inducción mostrando que la causalidad es anterior a la causación. La existencia de regularidades (causalidad) genera en nosotros la ilusión de la conexión necesaria, y esta ilusión nos da la convicción de que todo suceso tiene una causa (causación).

Por lo tanto, el legado de Hume sobre la causación implica al menos dos principios fundamentales. Primero, en la naturaleza no hay ninguna conexión necesaria. Y segundo, en ella encontramos regularidades universales en vez de conexiones causales. El escepticismo de Hume con respecto a la conexión necesaria no lo conduce a negar la existencia de toda verdad en la causación. Antes bien, hay una verdad, pero no la esperada. Esperábamos que hubiera un vínculo causal entre la causa y el efecto, pero lo que encontramos es, de hecho,

una secuencia de sucesos que ejemplifican las leyes universales. Estos dos aspectos han ejercido su influencia sobre el debate de la causación hasta nuestros días. La mayor parte de los filósofos cree que en la naturaleza no hay conexiones causales y que cualquier conexión causal específica debe ser el ejemplo de una ley universal. La mayoría se empeña en señalar que los términos utilizados para formular la ley no deben ser necesariamente iguales a los términos por los cuales se describen los incidentes de la relación causal original. Así, si digo: "Lo que John hizo causó el fenómeno visto por Sally", y supongo que John puso la olla con agua en la cocina y encendió el fuego, y Sally vio agua hirviendo en la olla, sería cierto que el acto de John causó el fenómeno visto por Sally, pero no habría ninguna ley que mencionara a John y Sally y ni siquiera los actos de poner y ver. Las leyes científicas se referirán a cosas como la presión del agua cuando esta se calienta en la atmósfera terrestre.

El escepticismo de Hume con respecto a la inducción ha tenido menos influencia en la filosofía contemporánea que su teoría de la regularidad de la causación. A mi juicio, la mayor parte de los filósofos de nuestros días creen poder responderle; la respuesta convencional dada por los manuales es que Hume se equivocó al suponer que los argumentos inductivos debían satisfacer criterios deductivos. Nuestro filósofo estima que algo falta en un argumento que procede mediante métodos inductivos sobre la base de evidencias para respaldar una conclusión, porque las premisas no entrañan esta última a la manera del argumento deductivo. Según el punto de vista de los filósofos contemporáneos, es como si alguien dijera: "Mi motocicleta no es buena porque no obtiene buenas calificaciones en una expo-

sición canina". Las motocicletas no son lo mismo que los perros ni se las debe juzgar de acuerdo con los criterios aplicados a estos. Se comete exactamente la misma clase de error cuando se supone que los argumentos inductivos deben juzgarse mediante criterios deductivos. A través de estos se obtienen argumentos deductivos válidos, y mediante los criterios inductivos hay argumentos inductivos válidos. Es un error confundir unos con otros.

En rigor, según una visión convencional contemporánea, aun esto implica conceder demasiado a Hume. La idea misma de que hay dos estilos de argumentos, inducción y deducción, ya es una fuente de confusión. Sólo hay argumentos deductivos, y una manera de proceder en las ciencias recibe el nombre de método hipotético deductivo. Uno formula una hipótesis, deduce una predicción y luego somete a prueba la primera viendo si la segunda resulta cierta. Cuando la predicción demuestra ser verdadera, decimos que la hipótesis original se confirma u obtiene respaldo. Cuando la predicción no se verifica, decimos que la hipótesis no se confirma o es refutada. No hay una oposición tajante entre inducción y deducción. Antes bien, la llamada inducción tiene que ver con la puesta a prueba de hipótesis mediante experimentos y otros tipos de evidencias. Y una manera típica de someter a prueba una hipótesis consiste en deducir sus consecuencias y luego ver si estas pueden pasar determinadas pruebas experimentales. Por ejemplo, la ley de la gravedad predice que un cuerpo caerá cierta distancia al cabo de cierto tiempo. Tras hacer esta deducción, sometemos a prueba la hipótesis viendo si los objetos caen efectivamente esa distancia en el lapso previsto.

II. ¿Nunca experimentamos la causación?

Dije antes que siento una gran admiración por los logros de Hume en su análisis de la conexión necesaria y su teoría de la regularidad de las relaciones causales. Pero también debo decir que la teoría me parece desastrosamente errónea y que tuvo un muy mal efecto sobre la filosofía ulterior. En este libro no voy a emprender una crítica general de la explicación de la causación y la inducción propuesta por Hume; sólo me concentraré en los rasgos esenciales para la filosofía de la mente. El principal resultado negativo de Hume en cuanto a la conexión necesaria puede enunciarse en una frase: no hay impresión de una conexión necesaria; es decir, no hay experiencia de la fuerza, la eficacia, el poder o la relación causal. ¿Es eso correcto? ¿Al lector le parece plausible? Debo confesar que a mí no me parece plausible en absoluto. Creo que a lo largo de nuestra vida despierta tenemos una percepción bastante grande de las conexiones necesarias, y quiero explicar cómo.

Cuando tenemos experiencias perceptivas o nos dedicamos a actos voluntarios, como vimos en nuestra discusión de la intencionalidad, hay una condición causalmente autorreferencial en las condiciones de satisfacción de los fenómenos intencionales. La intención en la acción sólo se cumple si causa el movimiento corporal, y la experiencia perceptiva sólo se lleva a cabo si es causada por el objeto percibido. Pero en ambos casos es muy común –aunque no, desde luego, de validez universal– que experimentemos efectivamente la conexión causal entre la experiencia, por un lado, y los objetos y situaciones del mundo, por otro. Si el lector tiene alguna duda acerca de esto, que levante el brazo. Es evidente que hay una distinción entre la experien-

cia de levantar el brazo y la de que sea otro quien nos lo levante. Como mencioné en el capítulo 5, el neurocirujano Wilder Penfield comprobó que podía mover el brazo de su paciente si estimulaba con microelectrodos las neuronas de la corteza motriz. Los pacientes decían invariablemente algo así como "yo no lo hice, fue usted"[2]. Ahora bien, como es obvio, esta experiencia es diferente de la de levantar real y voluntariamente el brazo. En el caso normal, cuando uno levanta el brazo adrede, experimenta concretamente la eficacia causal de la intención consciente en la acción que produce el movimiento corporal. Por otra parte, si alguien tropieza con nosotros, tenemos cierta percepción, pero no la experimentamos como si nosotros fuéramos su causa. Sentimos que ha sido efectivamente causada por el cuerpo de la persona que nos atropella. Así pues, en ambos casos, tanto en la acción como en la percepción, me parece muy común y hasta normal que percibamos una conexión causal entre objetos y situaciones del mundo y nuestras experiencias conscientes. En el caso de la acción sentimos que nuestras intenciones conscientes en la acción causan movimientos corporales. En el caso de la percepción sentimos que los objetos y situaciones del mundo causan experiencias perceptivas en nosotros.

A mi entender, Hume buscaba en el lugar equivocado. Lo hacía de una manera imparcial en objetos y sucesos fuera de sí mismo, y descubrió de ese modo que no había una conexión necesaria entre ellos. Pero si consideramos la índole de nuestras experiencias reales, es muy común sentir, me parece, que nosotros mismos hacemos suceder algo (esto es, una acción intencional)

2 W. Penfield, *The Mystery of the Mind*, *op. cit.*, p. 76.

o que algo hace suceder alguna otra cosa en nosotros (esto es, una percepción). En uno y otro caso es muy corriente experimentar la conexión causal.

Elizabeth Anscombe dio (en conferencias) un buen ejemplo de lo que decimos. Supongamos que estoy sentado tras mi escritorio y la detonación del escape de un automóvil afuera me hace dar un salto. En este caso siento en concreto que mi movimiento involuntario ha sido causado por el ruido fuerte que acabo de oír. No debo esperar la conjunción de instancias semejantes. Ahora experimento realmente el nexo causal como parte de mi secuencia de experiencias conscientes.

Hasta aquí, esas experiencias sólo nos darían una relación causal entre nuestras propias experiencias y el mundo real, pero me gustaría poder descubrir la misma relación en este último, al margen de aquellas. No me parece difícil en absoluto extender la concepción de la causación que sacamos de nuestras experiencias a los objetos y situaciones del mundo que existen e interactúan unos con otros, y hacerlo de una manera totalmente independiente de dichas experiencias. El efecto que yo mismo creo cuando causo el movimiento del automóvil al empujarlo es un efecto que puedo notar cuando te observo mientras lo empujas. Pero la relación causal es la misma, con prescindencia de que yo lo empuje o te vea hacerlo. Por otra parte, puedo ampliar esta anotación al caso en que no participa ningún agente. Si veo un auto que empuja otro, veo la fuerza física del primero como causante del movimiento del segundo. Parece entonces que, además de nuestras experiencias reales de causación, podemos extender con facilidad esta noción a secuencias de sucesos del mundo que no contienen dichas experiencias ni, para el caso, las de ninguna otra persona. Después de todo, las

relaciones causales con participación de seres humanos son sólo una parte ínfima de las relaciones causales del universo. El quid para la presente discusión es que la existencia de la misma relación que experimentamos cuando hacemos suceder algo o cuando algo hace suceder alguna otra cosa en nosotros puede percibirse aunque la relación causal no implique ninguna experiencia.

Nuestra experiencia de la causación no es por sí misma garantía de nada. Podríamos estar equivocados en cualquier caso específico. Pero esta posibilidad de error e ilusión está incluida en toda experiencia perceptiva. Lo importante en este análisis es destacar que la experiencia de la causación no es peor que cualquier otra experiencia perceptiva.

III. La causación mental y el cierre causal de lo físico

Supongamos que hasta aquí tengo razón: que, en efecto, tenemos la experiencia de la causación como parte de nuestra conciencia despierta normal, y que la causación es una relación real en el mundo real. De todas maneras, la causación mental parece presentar un problema especial, a saber: si la conciencia no es física, ¿cómo puede llegar a tener un efecto físico, como el de mover mi cuerpo? No obstante, nuestra experiencia nos dice, al parecer, que la conciencia lo mueve. Tomo la decisión consciente de levantar el brazo y el brazo se levanta. Al mismo tiempo, sin embargo, sabemos que puede contarse otra historia sobre el brazo que se levanta, una historia vinculada con las activaciones neuronales en la corteza motriz, la secreción de acetilcolina en las placas terminales de los axones de mis

neuronas motrices, la estimulación de los canales iónicos y el ataque al citoplasma de la fibra muscular, hasta que finalmente el brazo se alza. Así pues, si debe contarse una historia sobre el efecto de la conciencia en el nivel de la mente, ¿cómo casa con el relato que es preciso contar acerca de la química y la fisiología en el nivel del cuerpo? Peor aún, si suponemos que podemos asignar un papel a la causación mental y que la mente desempeña un papel causal en la producción de nuestro comportamiento corporal, será como salir de Guatemala para entrar a Guatepeor, porque ahora tenemos demasiadas causas. Al parecer, estamos ante lo que los filósofos llaman "sobredeterminación causal". Habría dos series independientes de causas que motivan el levantamiento de mi brazo, una relacionada con las neuronas y otra vinculada con la intencionalidad consciente.

Ahora podemos resumir con cierta precisión el problema filosófico de la causación mental: si los estados mentales son estados no físicos reales, cuesta entender cómo pueden tener algún efecto sobre el mundo físico. Pero si lo tienen, nos topamos con una sobredeterminación causal. De una u otra manera, al parecer no podemos dar un sentido a la idea de causación mental. Hay cuatro proposiciones que, en conjunto, son inconsistentes.

1. La distinción entre la mente y el cuerpo: lo mental y lo físico constituyen reinos diferentes.
2. El cierre causal de lo físico: el reino físico está causalmente cerrado, en cuanto ningún elemento no físico puede entrar a él y actuar como causa.
3. El principio de exclusión causal: cuando las cau-

sas físicas son suficientes para explicar un suceso, no puede haber ningún otro tipo de causas de este.

4. La eficacia causal de lo mental: los estados mentales funcionan realmente de manera causal[3].

Juntas, estas cuatro proposiciones son incompatibles. Una salida es renunciar a la cuarta, pero esto equivale a caer en el epifenomenalismo. Como escribe Jaegwon Kim: "Si esto es epifenomenalismo, saquémosle el máximo provecho"[4].

En general, como hemos visto una y otra vez, cuando creemos estar frente a uno de estos problemas filosóficos imposibles, la realidad es que hemos planteado un supuesto falso. Me parece que así sucede en el presente ejemplo. El error se expresa en la primera proposición, la tradicional distinción entre la mente y el cuerpo. Dije en el capítulo 4 que ese error obedece a suponer que si hay un nivel de descripción de los procesos cerebrales en el cual estos contienen secuencias reales e irreductibles de estados conscientes, y hay otro nivel de descripción de esos mismos procesos en el cual estos son fenómenos puramente biológicos y los estados de conciencia no se pueden reducir en términos ontológicos a los fenómenos neurobiológicos, los dos niveles deben tener existencias separadas. En el capítulo 4 vimos que esto es erróneo. La salida de este dilema pasa por recordar una conclusión a la que llegamos en ese capítulo: la realidad e irreductibilidad de la conciencia no implica que se trate de un tipo independiente de entidad o propiedad situada "por encima" del sistema cerebral, en el cual se realiza físicamente. En el cerebro, la conciencia no es una entidad o propiedad independiente: es sólo *el estado en que se encuentra el cerebro.*

3 J. Kim, *Mind in a Physical World...*, *op. cit.*

4 J. Kim, "Causality, Identity and Supervenience in the Mind-Body Problem", en P. A. French, T. E. Uehling, Jr. y H. K. Wettstein (comps.), *Studies in Metaphysics*, *op. cit.*, p. 47.

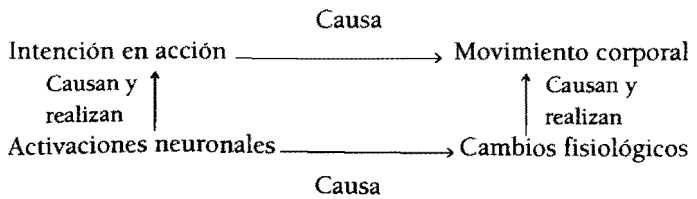
Nuestro vocabulario tradicional hace casi imposible formular este punto. Si decimos que lo mental es irreductible a lo físico, parecemos aceptar el dualismo. Pero si decimos que lo mental es simplemente lo físico en un nivel más elevado de descripción, admitimos en apariencia el materialismo. La salida, para insistir en un aspecto que planteé una y otra vez, consiste en abandonar el vocabulario tradicional de lo mental y lo físico y tratar de limitarse a enunciar los hechos. La relación de la conciencia con los procesos cerebrales es como la relación de la solidez del pistón con el comportamiento molecular de las aleaciones metálicas, de la liquidez de una extensión de agua con el comportamiento de las moléculas de H_2O, o de la explosión en los cilindros de un automóvil con la oxidación de las moléculas combustibles individuales. En todos los casos, las causas de nivel superior, en el plano sistémico global, no son algo adicional a las causas en el micronivel de los componentes del sistema. Antes bien, las causas de nivel sistémico son íntegramente explicadas por la causación de los microelementos y desde un punto de vista causal pueden reducirse por completo a ella. Esto es cierto tanto de los procesos cerebrales como de los motores de auto o del agua que circula en una lavadora. Cuando digo que mi decisión consciente de levantar el brazo hizo que este se levantara, no estoy diciendo que se presentó alguna causa *sumada* al comportamiento exhibido por las neuronas al activarse y producir toda clase

de consecuencias neurobiológicas; no hago sino describir simplemente la totalidad del sistema neurobiológico en su nivel de conjunto y no en el plano de microelementos específicos. La situación es el análogo exacto de la explosión en los cilindros del automóvil. Puedo decir que esa explosión causó el movimiento del pistón o bien que la oxidación de las moléculas combustibles liberó energía calórica y esta ejerció presión sobre la estructura molecular de las aleaciones. No se trata de dos descripciones independientes de dos conjuntos de causas independientes, sino de descripciones de un único sistema en dos niveles diferentes. Desde luego, como todas las analogías, esta funciona hasta cierto punto y nada más. La diferencia entre el cerebro y el motor de un automóvil radica en el hecho de que la conciencia no es ontológicamente reducible, como sí lo es la explosión en el cilindro a la oxidación de las moléculas individuales. Sin embargo, he sostenido antes y repetiré aquí lo siguiente: la irreductibilidad ontológica de la conciencia no proviene del hecho de que deba desempeñar un papel causal independiente; antes bien, se debe a que tiene una ontología de primera persona y, por ello, no es posible reducirla a algo con una ontología de tercera persona, aun cuando no hay una eficacia causal de la conciencia que no sea reducible a la eficacia causal de su base neuronal.

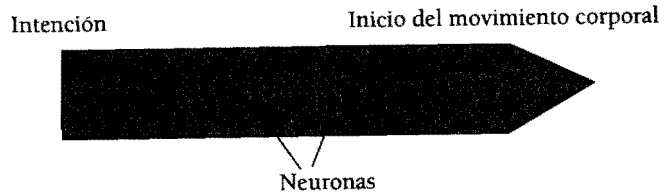
Podemos sintetizar de la siguiente manera el análisis desplegado en esta sección. Se supone que la causación mental plantea dos problemas: primero, ¿cómo puede lo mental, que es ingrávido y etéreo, afectar de algún modo el mundo físico? Y segundo, si lo mental funcionara causalmente, ¿no produciría una sobredeterminación causal? La forma de responder a estas preguntas consiste en dejar de lado los supuestos que, ante

todo, les dieron origen. El supuesto básico era que la irreductibilidad de lo mental implicaba su posición por encima de lo físico y no como una parte del mundo físico. Una vez que desechamos este supuesto, la respuesta a los dos enigmas es, en primer lugar, que lo mental es simplemente una característica (en el nivel del sistema) de la estructura física del cerebro, y en segundo lugar, que en términos causales no hay dos fenómenos independientes, el esfuerzo consciente y las activaciones neuronales inconscientes. Sólo está el sistema cerebral, que tiene un nivel de descripción en el cual ocurren las activaciones neuronales y otro nivel de descripción, el del sistema, en el cual este es consciente e intenta conscientemente levantar el brazo. Cuando abandonamos las categorías cartesianas tradicionales de lo mental y lo físico y renunciamos a la idea de la existencia de dos reinos desconectados, la causación mental no plantea, en realidad, ningún problema especial. Hay, desde luego, problemas muy arduos con respecto a su funcionamiento concreto en la neurobiología, cuyas soluciones aún no conocemos en su mayor parte.

Una manera de representar la relación es presentarla en un diagrama como el siguiente, donde el nivel superior muestra la intención en acción causando un movimiento corporal, mientras que el nivel inferior muestra su funcionamiento en el aparato neuronal y fisiológico. En cada paso, el nivel inferior causa y realiza el nivel superior:



Aunque pedagógicamente útiles, estos diagramas pueden ser engañosos si sugieren que el nivel mental está arriba, como la cobertura de una torta. Tal vez sea mejor proponer otra representación gráfica en la cual la intención consciente se muestre como existente en todo el sistema y no sólo en la parte de arriba. En el diagrama siguiente, los círculos representan las neuronas y el sombreado representa el estado consciente tal como se difunde por todo el sistema neuronal:



IV. La causación mental y la explicación del comportamiento humano

A lo largo de este libro hemos visto que hay dos tipos un tanto diferentes de problemas filosóficos en torno de los tópicos de la filosofía de la mente. Por un lado tenemos los problemas tradicionales de la forma: ¿cómo es posible tal cosa? Por ejemplo, cómo es posible que los estados cerebrales causen la conciencia. Pero también hay interrogantes de la siguiente forma: ¿cómo funciona en la vida real? ¿Cuáles son la estructura y función concretas de la conciencia humana? En este capítulo hemos examinado precisamente esa distinción entre la pregunta "¿cómo es posible que haya causación mental?" y la pregunta "¿cómo funciona en la vida real?" Quiero terminar diciendo al menos algo sobre el funcionamiento de la causación mental en la vida real. La comprensión de la respuesta a esta pregunta es absolutamente esencial para entendernos como seres humanos, pues cuando encaramos acciones voluntarias solemos hacerlo sobre la base de *razones*, y estas actúan de manera causal para explicar nuestro comportamiento; sin embargo, la forma lógica de la explicación del comportamiento humano en términos de razones es radicalmente diferente de las formas convencionales de causación. Quiero explicar ahora algunas de las diferencias.

En un caso típico de la causación no mental corriente decimos cosas como esta: "El derrumbe de la autopista fue causado por el terremoto". Pero si contrastamos esta afirmación con una explicación que solemos dar de nuestros propios actos (y siempre es una buena idea considerar nuestro caso, para ver cómo funciona la causación intencional en nuestra vida), veremos que la estructura lógica de esta última es radicalmente diferente. Supongamos que digo: "En las últimas elecciones voté por Bush porque quería una mejor política educacional".

Si se observa la primera explicación, sobre el derrumbe de la autopista, se verá que tiene varios rasgos lógicos interesantes. Primero, la causa enuncia una condición suficiente para la ocurrencia del efecto en ese contexto. Esto es: en ese contexto específico, dadas la estructura de la autopista y las fuerzas generadas por el terremoto, una vez que este ocurrió la autopista debía derrumbarse. Segundo, no hay finalidades ni metas en cuestión: el terremoto y el derrumbe son meros sucesos que ocurren. Tercero, aunque la explicación, como cualquier acto de habla, tiene un contenido intencional, este mismo no funciona de manera causal: el contenido intencional "terremoto" o "hubo un terremoto" no hace sino describir un fenómeno y no es cau-

sante de nada. Ahora bien, estas tres condiciones están ausentes en la explicación de mi comportamiento electoral. En mi caso, la explicación no ha enunciado condiciones suficientes. Sí, yo quería una mejora en la educación; sí, creí que Bush haría más que Gore por la educación, pero de todos modos, nada me obligaba a votar como lo hice. Podría haber votado por el otro candidato, en igualdad de las restantes condiciones. Segundo, no entenderemos la explicación a menos que veamos que se la enuncia en términos de las metas del agente. Las nociones de metas, objetivos, finalidades, teleología, etc., tienen una intervención esencial en este tipo de explicación. En rigor, mi explicación concreta es incompleta. Sólo entenderemos la afirmación de que un agente hizo A porque quería conseguir B si suponemos que el agente también *creía* que al hacer A produciría B, o al menos haría más probable su ocurrencia. Y tercero, en lo concerniente a estas explicaciones en términos de causación intencional es absolutamente esencial entender que el contenido intencional presente en ellas, por ejemplo que yo quería una mejor política educacional, aparece realmente en la causa misma cuya especificación dilucida el comportamiento que tratamos de explicar.

Estas tres características –el supuesto previo de la libertad, la exigencia de que una explicación de la acción especifique una meta u otro motivador y el funcionamiento de la causación intencional como parte del mecanismo explicativo– son muy diferentes de los componentes de las explicaciones de fenómenos naturales como los terremotos y los incendios forestales. Las tres forman parte de un fenómeno mucho más amplio, la racionalidad. Es esencial ver que el funcionamiento de la intencionalidad humana exige la presencia de la

racionalidad como principio organizador estructural y constitutivo de la totalidad del sistema. No puedo exagerar la importancia de este fenómeno para la comprensión de las diferencias entre las explicaciones naturalistas que nos dan las ciencias naturales y las explicaciones intencionalistas propuestas por las ciencias sociales. En la estructura superficial de las frases las siguientes explicaciones se parecen mucho:

1. Hice una marca en la boleta electoral porque quería votar por Bush.
2. Me dio dolor de estómago porque quería votar por Bush.

Aunque la estructura superficial es similar, la forma lógica real es muy diferente. La segunda frase sólo enuncia que un suceso, mi dolor de estómago, fue causado por un estado intencional, mi deseo. La primera, en cambio, no enuncia una condición causalmente suficiente y sólo tiene sentido dentro del contexto de una teleología supuesta por anticipado.

Estas explicaciones plantean una multitud de problemas filosóficos. El más importante es el problema del libre albedrío, del que me ocuparé en el próximo capítulo.

CAPÍTULO 8

EL LIBRE ALBEDRÍO

Los problemas filosóficos tienden a agruparse. Para resolver e incluso abordar uno de ellos, por lo común es preciso ocuparse de varios otros. El problema del libre albedrío es un ejemplo particularmente llamativo de este fenómeno general. Para abordarlo, tenemos que explorar la naturaleza de la conciencia, la causación, la explicación científica y la racionalidad. Peor aún, luego de examinar todas esas cuestiones y su relación con el problema del libre albedrío, habremos aclarado este pero todavía careceremos de una solución; por lo menos, yo no soy capaz de divisar un camino para llegar a ella. Mi única expectativa real en este capítulo es explicar cuáles son las cuestiones y cuáles serían las posibles soluciones. La conclusión general a la que llego es que necesitaremos saber mucho más sobre las operaciones del cerebro antes de alcanzar una solución al problema del libre albedrío de cuya pertinencia podamos estar seguros.

I. ¿Por la libertad de la voluntad es un problema?

En lo concerniente al libre albedrío hay una situación especial, porque tenemos dos convicciones absolutamente inconciliables, pero ambas parecen del todo acertadas y hasta ineludibles. La primera es que todo hecho ocurrido en el mundo tiene causas suficientes antecedentes. Las causas suficientes de un hecho son aquellas que, en un contexto específico, bastan para

determinar su ocurrencia. Cuando decimos que las causas fueron suficientes queremos decir que, visto que se produjeron en ese contexto histórico, el suceso mismo tenía que ocurrir. Cuando pedimos una explicación de un suceso, no nos satisfacen las razones que se limitan a presentarlo como parte de una secuencia de acontecimientos. Queremos saber qué hizo que el hecho sucediera. Queremos saber por qué ocurrió ese hecho y no otros que podrían haber acontecido. La imagen que tenemos es que todos los sucesos del mundo están tan determinados como, por ejemplo, la caída de esta pluma en caso de soltarla. Si suelto la pluma que tengo en la mano, en este contexto caerá sobre la mesa. Dada la estructura del universo, si la suelto tiene que caer encima de la mesa porque las fuerzas que actúan sobre ella son causalmente suficientes para determinar esa caída. Nuestra convicción determinista equivale a la idea de que lo valedero para la caída de la pluma es valedero para cualquier hecho que haya sucedido o vaya a suceder.

Nuestra segunda convicción, a saber, que en realidad tenemos libre albedrío, se basa en ciertas experiencias de libertad humana. Vivimos la experiencia de decidirnos a hacer algo y luego hacerlo. Como parte de nuestras experiencias conscientes, sentimos que las causas de nuestras decisiones y acciones, en forma de razones para unas y otras, no son suficientes para forzar las decisiones y acciones concretas. Piénsese en lo que implica decidir por qué candidato votar en unos comicios, e incluso la elección de un plato en el menú de un restaurante, y se advertirá que la toma de decisiones entraña una experiencia característica, parte de cuyo contenido es el hecho de tener una idea de las alternativas a nuestra disposición. En síntesis, hay una

distancia entre las causas de nuestras decisiones y acciones en forma de razones y la toma efectiva de las primeras y la ejecución de las segundas. La decisión y la acción voluntarias contrastan con la percepción en cuanto hay en ellas un intervalo entre las causas del fenómeno, con forma de razones para la decisión o la acción, y la aparición real de una u otra, mientras que en la percepción ese intervalo no existe. Por eso existe el problema de "la libertad de la voluntad", pero no el problema de "la libertad de la percepción". Si me miro la mano puesta a la altura de la cara, las causas, esto es, que tengo la mano directamente frente a los ojos abiertos, la luz es adecuada y mis ojos están sanos, son suficientes para producir la experiencia visual. No hay intervalo. En las acciones voluntarias, en cambio, hay al menos tres intervalos o, para decirlo con mayor precisión, al menos tres fases de un intervalo continuo. Hay un intervalo entre el conocimiento de las razones para la acción y la decisión de llevarla a cabo. Por ejemplo, en un caso típico en el que se nos pide que elijamos entre Smith y Jones, dos candidatos en una elección, el conjunto de razones que tenemos para votar por uno u otro no suele forzar por sí mismo nuestra decisión. Segundo, hay un intervalo entre la decisión y la iniciación efectiva de la acción. Por ejemplo, una vez que hemos resuelto votar por Jones, la decisión no fuerza la acción. Al entrar al cuarto oscuro aún nos resta *actuar* en consonancia con ella. Tercero, para cualquier serie extensa de acciones, por ejemplo mi intento de aprender ruso o de escribir un libro sobre filosofía de la mente, hay un intervalo entre el comienzo de la acción y su continuación hasta completarse. Uno no puede, por decirlo así, darse un empujón y dejar que su movimiento prosiga como un tren que se desplaza sobre las vías.

No, debe hacer un esfuerzo constante para seguir adelante con la acción hasta su término.

Ahora bien, dicho todo esto, tengo que hacer de inmediato algunas salvedades. A veces hay un intervalo en la percepción, como ocurre por ejemplo cuando pasamos de ver una figura esbozada como un pato a verla como un conejo. Pero en realidad esto no contradice el argumento general, porque en estos casos hay un elemento voluntario en la percepción. Depende de nosotros ver la figura como un pato o como un conejo. Y no todas las acciones humanas, por supuesto, contienen una experiencia del intervalo. A menudo nos sentimos en las garras de un impulso o una emoción avasallantes, en cuyo caso estamos impedidos de ver posibilidades alternativas. Pero ése es precisamente el contraste entre las acciones voluntarias libres, por un lado, y las acciones compulsivas, adictivas u obsesivas, por otro.

Nuestra experiencia del intervalo es la base de la convicción de que tenemos libre albedrío. Pero ¿por qué debemos apreciar tanto esas experiencias? Después de todo, tenemos muchas experiencias que sabemos ilusorias. ¿Por qué no aceptar simplemente que el libre albedrío es una ilusión como lo es, por ejemplo, el color a juicio de algunos filósofos? No se trata, empero, de una experiencia que podamos desestimar con ligereza adjudicándole ese carácter de mera ilusión. Cada vez que tomamos una decisión, debemos presuponer la libertad. Por ejemplo, si estoy en un restaurante con un menú en la mano y el camarero me pregunta qué voy a pedir, no puedo decirle: "soy determinista, sencillamente esperaré hasta ver qué pasa", porque aun ese enunciado sólo es inteligible para mí como un ejercicio de mi libre albedrío. No puedo concebirlo como algo que simplemente me ocurrió, a la manera de un dolor

repentino en el estómago. Una de las curiosidades de la experiencia del libre albedrío es que no podemos deshacernos de la convicción de que somos libres, aun cuando estemos filosóficamente persuadidos de que esa convicción es errónea. Cada vez que decidimos o actuamos de manera voluntaria, cosa que hacemos a lo largo del día, debemos decidir o actuar sobre la base del supuesto previo de nuestra libertad. De lo contrario, nuestras decisiones y acciones nos resultan ininteligibles. No podemos apartar nuestro libre albedrío del pensamiento.

Al parecer, entonces, debemos tener, por una parte, la profunda convicción de que todo hecho que ocurre debe explicarse a través de condiciones causalmente suficientes, y por otra, las experiencias que nos dan la convicción de la libertad humana, una convicción que en la práctica no podemos abandonar, por mucho que reneguemos de ella en la teoría.

II. ¿Es el compatibilismo una solución al problema del libre albedrío?

Creo que la mayoría de los filósofos aceptan hoy de uno u otro modo la idea de que, si entendemos estas nociones como corresponde, podremos ver que la tesis del libre albedrío es en realidad compatible con la tesis del determinismo. Tanto el determinismo como el libre albedrío son verdaderos. Poco sorprenderá al lector saber que esta concepción se denomina "compatibilismo"; en su origen fue bautizada por William James como "determinismo blando", para contrastarla con el "determinismo duro", la tesis de que el libre albedrío y el determinismo son incompatibles, porque este es verdadero y aquel, falso. De acuerdo con los compati-

bilistas, decir que una acción es libre no significa decir que carece de condiciones antecedentes causalmente suficientes, sino que tiene algún tipo de condiciones causales. Así, por ejemplo, si ahora decido levantar el brazo derecho y lo hago, en esas condiciones lo levanto en virtud de mi libre albedrío; y en términos más grandilocuentes, si decido escribir la gran novela norteamericana o votar por el candidato republicano, se trata también de decisiones que tomo y llevo a la práctica en función de mi libre albedrío. Ahora bien, los compatibilistas sostienen, desde luego, que tienen causas como todo lo demás, que están íntegramente determinadas desde el punto de vista causal. El quid, empero, es que la determinación procede de mis convicciones, procesos racionales y reflexiones internas. De modo que las acciones libres no son acciones indeterminadas; están tan determinadas como cualquier otro hecho que ocurre en el mundo. Pero su libertad consiste en estar determinadas por cierto tipo de causas y no por otras. Pongamos por caso: si decido levantar el brazo a fin de dar un ejemplo filosófico, se trata de una acción voluntaria libre. Pero si un hombre me pone un revólver en la cabeza y me dice: "¡Levante el brazo derecho!", cuando lo hago no actúo libremente. Actúo bajo amenaza, fuerza o compulsión. "Libre", en resumen, no se contrapone a "causado" sino a "forzado", "obligado", "bajo coacción", etcétera.

Al parecer, según la visión compatibilista podemos repicar y estar en la procesión. Podemos decir: sí, todas las acciones están determinadas, pero algunas son libres porque la determinación proviene de cierto tipo de procesos psicológicos internos, formas de racionalidad, deliberación, etcétera.

¿Nos da el compatibilismo una verdadera solución al problema del libre albedrío? Dije que, a mi entender, la mayoría de los filósofos estiman que sí. Y el compatibilismo tiene, por cierto, una larga y distinguida historia. En diferentes versiones adhirieron a él Thomas Hobbes, David Hume, John Stuart Mill y, en el siglo xx, A. J. Ayer y Charles Stevenson. La convicción de que esta doctrina brinda una solución al problema del libre albedrío dependerá de nuestra visión de este. Si el problema se refiere al uso corriente de expresiones como "por mi libre voluntad", parece claro que hay una utilización en la que el hecho de decir que actué por mi libre voluntad deja abierta la cuestión de si las causas antecedentes fueron causalmente suficientes. Hay un uso de las palabras, en efecto, que es consistente con el compatibilismo, pero no se trata del problema original del libre albedrío que nos preocupaba. Cuando la gente marcha por las calles llevando carteles que exigen "Libertad ya", por lo común no piensa en la naturaleza de la causación; sólo quiere que el gobierno la deje en paz o algo parecido. Y ése es, sin duda, un uso importante del concepto de libertad, pero no es el concepto central para el problema del libre albedrío; no lo es, al menos, según mi interpretación de este. El problema es el siguiente: ¿son todas nuestras decisiones y acciones precedidas por condiciones causalmente suficientes, condiciones que bastan para determinar la ocurrencia de dichas decisiones y acciones? ¿La secuencia del comportamiento racional humano y animal está determinada como lo está en su movimiento la caída de la pluma sobre la mesa, por la fuerza de gravedad y otras fuerzas que actúan sobre ella? El compatibilismo no da respuesta a estas preguntas.

El compatibilismo hace un planteamiento lógico sobre los conceptos de "libre" y "determinado" y señala, con acierto, que hay un uso de estos según el cual decir que una acción es libre no significa, hasta aquí, formular interrogante alguno acerca de si está determinada o no, en el sentido de la existencia de condiciones causales previamente suficientes. Pero una vez aceptado ese planteamiento lógico, todavía queda abierta una cuestión empírica fáctica. *¿Es verdad que toda acción humana ocurrida en el pasado, que ocurre ahora o que ocurrirá alguna vez fue, es y será causada por condiciones previamente suficientes?* ¿Son las causas de todas nuestras acciones condiciones causales suficientes? Admitiendo que nuestras acciones tienen causas y que algunas de ellas, como las de características compulsivas, son causadas por condiciones previamente suficientes, ¿las causas de cualquier acción bastan para determinar que debe ocurrir esa acción y ninguna otra? El compatibilismo no responde y ni siquiera aborda este problema del libre albedrío. La teoría supone simplemente que estamos determinados. Pero la cuestión sigue abierta después de aceptar el planteamiento compatibilista sobre ciertos usos lingüísticos. Adviértase que la cuestión del libre albedrío, tal como la he enunciado, no se vale de manera esencial de nociones como "libertad", "por mi propia libre voluntad", "voluntario", etc. Sólo se refiere a condiciones causalmente suficientes.

Creo que otro motivo por el cual muchos filósofos aceptan el compatibilismo es que en realidad no están muy interesados en el problema del libre albedrío tal como yo lo he definido. Se interesan en el problema de la "responsabilidad moral". Se afanan en insistir en que una persona como Hitler no escapa a la responsabilidad moral por sus acciones aun cuando podamos mostrar

que su comportamiento estaba determinado. En ese sentido, quieren afirmar que la responsabilidad moral es compatible con el determinismo; y como al menos en un sentido de "libre" parece haber una conexión entre esa responsabilidad y la libertad, habría que deducir que debe haber un sentido de esa misma palabra compatible con el determinismo. Aunque interesantes, estos temas no corresponden a mis inquietudes en el presente libro. Mi problema puede enunciarse con independencia de todas esas disputas acerca del determinismo y la responsabilidad moral. La cuestión, repitámoslo, es si para toda acción humana (incluido el acto de decidir) que se haya realizado o se realizare alguna vez, podría haber causas antecedentes suficientes para determinar esa acción y ninguna otra.

Así, queda una cuestión fáctica: ¿qué es verdad, el determinismo o su negación (llamémosla "libertarianismo")? La cuestión tiene dos aspectos, uno psicológico y otro neurobiológico. Considerémoslos en orden.

III. ¿Es verdadero el determinismo psicológico?

El interrogante con respecto al determinismo psicológico es si nuestros estados psicológicos son causalmente suficientes para determinar todas nuestras acciones voluntarias. ¿Son nuestros estados psicológicos, en forma de creencias y deseos, esperanzas y temores, así como el conocimiento de nuestras obligaciones y compromisos, etc., causalmente suficientes para determinar todas nuestras decisiones y acciones? Nótese que la planteo como una cuestión empírica directamente fáctica. Lo primero que debe advertirse es que nuestra comprensión de esos conceptos se basa en el conoci-

miento de un contraste entre los casos en los cuales estamos auténticamente sometidos a compulsiones psicológicas y los casos en que no sucede así. El drogadicto, el alcohólico y otros compulsivos no tienen libertad psicológica. Dada la situación psicológica en que se encuentran, no son capaces de ayudarse a sí mismos. La pregunta, entonces, es: ¿todas las causas psicológicas son así? ¿Mi decisión de votar por el candidato republicano es exactamente igual a la conducta del drogadicto que toma heroína de manera compulsiva como resultado de su adicción?

Bien, argumentemos con el mayor rigor posible a favor de la tesis determinista. Hay muchos experimentos para mostrar que a menudo nos encontramos en una situación en la cual creemos comportarnos con libertad desde un punto de vista psicológico, pero en realidad nuestro comportamiento está determinado. Los más impresionantes tal vez sean los experimentos de hipnosis. En un ejemplo característico (un caso real), se pidió al sujeto que, una vez salido del trance hipnótico, al escuchar la palabra "Alemania" fuera hacia la ventana y la abriera. Tan pronto como escuchó esa palabra, el sujeto inventó un motivo de apariencia perfectamente racional para abrir la ventana. Para justificar su acción, dijo algo así: "El ambiente está terriblemente sofocante, necesitamos aire fresco. ¿No les molesta que abra la ventana?" Para él, la acción era completamente libre. Pero tenemos buenas razones para suponer que estaba determinada por causas desconocidas para el sujeto. En este caso, entonces, el intervalo es una ilusión. El sujeto creía de manera ilusoria llevar a cabo una acción libre, pero en realidad su comportamiento estaba totalmente determinado. Ahora, nuestra pregunta es la siguiente: ¿parece razonable suponer que todas las acciones tienen esas características? Bien, esta es una afirmación fáctica, que la reflexión filosófica no puede resolver. Sin embargo, parece muy improbable que todas nuestras acciones se realicen según el modelo del drogadicto o la persona que sale de un trance hipnótico. En este momento no estoy bajo hipnosis y, a decir verdad, nunca lo estuve. Si ahora decido qué voy a comer en el almuerzo o dónde pasaré la tarde, las causas psicológicas actuantes en mí son muy diferentes de las causas psicológicas que intervienen en el adicto o el sujeto luego de la hipnosis. He expuesto esos dos casos, el hipnotismo y la adicción, como si fueran iguales, pero en realidad creo que exhiben importantes diferencias. El hombre hipnotizado actúa en el intervalo, pero no conoce todas sus motivaciones. Tiene una motivación preponderante que ignora por completo. De hecho, psicológicamente hablando lleva a cabo una acción libre, pero su motivación preeminente es inconsciente. La libertad plena exige tener conocimiento de las propias motivaciones, cosa que no sucede con este agente. Hay aquí una diferencia con el adicto, que puede muy bien saberse preso de una adicción, no obstante lo cual se comporta de manera adictiva.

Hay una multitud de experimentos, similares a los de la hipnosis, en los cuales la gente vive una experiencia del intervalo, pero tenemos motivos independientes para creer que no son libres. Muchos científicos estiman que esos experimentos dan crédito a la hipótesis de que todas nuestras acciones están psicológicamente determinadas[1]. Pero en mi opinión tienden a respaldar la hipótesis contraria. Entendemos todos estos casos,

1 D. N. Wegner, *The Illusion of Conscious Will*, Cambridge, Mass., MIT Press, 2003.

de hipnosis, engaño, confabulación, etc., si los contrastamos con el caso convencional en el cual realizamos una acción voluntaria libre. Los casos en que el intervalo es una ilusión son justamente casos que difieren en algunos aspectos importantes de los ejemplos clásicos de acciones voluntarias. Creo entonces que no confirman por sí mismos el determinismo psicológico. Sin embargo, insistamos, el interrogante de si todas nuestras acciones están o no psicológicamente determinadas es una cuestión empírica fáctica, que la argumentación filosófica no puede zanjar por sí sola. Lo que sostengo en este momento es que las pruebas disponibles respaldan la idea de que tenemos libertad psicológica. Aun los casos en que esta falta se entienden en contraste con los casos en que está presente.

IV. ¿Es verdadero el determinismo neurobiológico?

A los efectos de seguir adelante con este capítulo, voy a aceptar la conclusión de que la libertad psicológica es real. Las causas puramente psicológicas de nuestras acciones no suelen ser suficientes en términos causales para determinar estas últimas. Sin embargo, esta constatación no resuelve un profundo interrogante: ¿qué pasa con la neurobiología subyacente? Podríamos tener libre albedrío en el nivel psicológico, esto es: la psicología no sería como tal suficiente para fijar nuestras acciones. Pero la neurobiología subyacente, que también determina esa psicología, podría ser causalmente suficiente para determinarlas. A lo largo de este libro hemos supuesto que en cualquier momento dado el estado de conciencia de una persona es causalmente determinado en su totalidad por su neurobiología.

Ahora sostenemos que los estados conscientes no suelen bastar para determinar las decisiones y acciones. Pero la cuestión sigue en pie: ¿es la neurobiología suficiente para determinarlas? Ocupémonos de este aspecto, que es a mi juicio el más serio del problema del libre albedrío.

Nos acercamos ahora al meollo del asunto, y por consiguiente es una buena idea recapitular para ver hasta dónde hemos llegado. En este capítulo y los anteriores establecí, o al menos propuse argumentos para establecer las siguientes conclusiones:

1. El libertarianismo psicológico, tal como lo he definido, es probablemente verdadero. La tesis dice que nuestros estados psicológicos, creencias, deseos, esperanzas, temores, etc., no son en todos los casos causalmente suficientes para determinar la acción ulterior. En lo que respecta al nivel psicológico, las acciones libres sin duda existen, aunque no todas las acciones, desde luego, son libres en ese plano. A veces, por ejemplo en los casos de compulsión, furia, deseo avasallante, etc., el agente cae en las garras de condiciones psicológicamente suficientes. Pero la presente discusión sostiene, entre otras cosas, que no todos los casos son así. Lo cual es sólo otra manera de decir que el intervalo es *psicológicamente* real y no ilusorio.
2. En capítulos anteriores afirmé que todos nuestros estados psicológicos sin excepción están, en cualquier instante dado, íntegramente determinados por el estado del cerebro en ese momento. Así, por ejemplo, en este preciso momento todos mis estados psicológicos, conscientes e inconscientes, están determinados por las actividades que se

desarrollan en el cerebro. Cualquier cambio en el estado psicológico exigiría un cambio en la actividad cerebral. Este punto nos permitió resolver el problema del epifenomenalismo. Nuestros estados conscientes son rasgos sistémicos o de nivel superior del cerebro y, por consiguiente, no constituyen dos conjuntos independientes de causas, las psicológicas y las neurobiológicas. Lo psicológico no es más que lo neurobiológico descrito en un nivel más elevado.

Pero si la libertad psicológica, la existencia del intervalo, marca una diferencia para el mundo, debe manifestarse de una manera u otra en la neurobiología. ¿Cómo puede hacerlo? Ya hemos visto que la neurobiología es en cualquier momento dado suficiente para fijar el estado total de la psicología en ese mismo instante, en virtud de una causación de abajo arriba. Así, la ausencia de condiciones causalmente suficientes en el nivel psicológico, la ausencia de condiciones suficientes en la causación psicológica que, por así decirlo, va de izquierda a derecha a través del tiempo, sólo significará una diferencia real si se refleja de alguna manera en el nivel neurobiológico. *Si la libertad es real, el intervalo debe descender hasta el nivel de la neurobiología.* Pero ¿cómo podría hacerlo? En el cerebro no hay intervalos.

V. Construcción de un caso de prueba

A fin de examinar esta cuestión, tendremos que construir un ejemplo en el que haya una clara diferencia fáctica entre una acción libre y una acción determinada. ¿Cómo sería exactamente el mundo si el

determinismo fuera cierto? ¿Cómo contrastaría con un mundo donde el libertarianismo fuera verdadero? Construyamos un ejemplo que ilustre la diferencia. Tomaremos un caso célebre, aunque mitológico. Zeus pidió a Paris, hijo del rey de Troya, que obsequiara una manzana de oro con la inscripción "a la más bella" a una de estas tres diosas: Afrodita, Palas Atenea y Hera. En contra de un malentendido corriente representado en muchas pinturas famosas, Paris no elegiría a la diosa de mejor apariencia sino a la que le ofreciera el soborno más suculento. Palas Atenea le propuso hacerlo gobernante de Europa y Asia. Hera se ofreció a permitirle conducir a los troyanos a la victoria militar sobre los griegos, y Afrodita se comprometió a entregarle la mujer más hermosa del mundo. Todos sabemos que Paris eligió a Afrodita, con consecuencias que no puede sino calificarse de desastrosas.

Armemos ahora el caso. Supondremos que en el momento t_1 Paris se enfrenta a la elección. Supondremos también que el estado total de su cerebro en el momento t_1 incluye un completo conocimiento de la elección, así como sus razones para tomar la decisión que fuere. En el momento t_2, digamos diez segundos después, el joven decide dar la manzana a Afrodita y su brazo se mueve con ese fin. Supongamos que en los diez segundos transcurridos entre t_1 y t_2 no entra en el cerebro de Paris absolutamente ningún estímulo exterior. Podemos imaginar que cierra los ojos, no oye nada y ningún estímulo externo vinculado a la decisión le llega al cerebro. La cuestión del libre albedrío puede enunciarse ahora con cierta precisión: si el estado total de su cerebro en t_1 es causalmente suficiente para determinar el estado total de su cerebro en t_2, en t_1 su decisión está completamente determinada. ¿Por qué?

Porque en t_2 Paris toma su decisión, y cuando la acetilcolina llega a las placas terminales axónicas de sus neuronas motrices, el brazo comienza a tenderse hacia Afrodita por una necesidad causal. Si el estado total del cerebro en t_1 es suficiente para fijar su estado total en t_2, en este caso y en todos aquellos que exhiban una similitud importante, ni Paris ni ninguno de nosotros tiene libre albedrío. Si funciona de ese modo en el aparato, por decirlo así, el libre albedrío es una masiva ilusión. Por el contrario, si el estado del cerebro en t_1 no es causalmente suficiente para fijar su estado en t_2, entonces, *dados ciertos supuestos cruciales sobre el papel de la conciencia*, el libre albedrío es una realidad.

Exploremos en orden cada una de las posibilidades.

Hipótesis 1: el determinismo y el cerebro mecánico

De acuerdo con la primera hipótesis debemos suponer que el cerebro es una máquina en el sentido tradicional y anticuado de los motores de automóvil, los motores de vapor y los generadores eléctricos. Se trata de un sistema completamente determinista y cualquier apariencia de indeterminismo es una ilusión basada en nuestra ignorancia, de modo que esta hipótesis se ajusta bien a lo que tendemos a creer de la naturaleza y la biología en general. El cerebro es un órgano como cualquier otro y no tiene más libre albedrío que el corazón, el hígado o el pulgar izquierdo. Esto también se adecua a una concepción vigente en la ciencia cognitiva, según la cual debemos imaginar el cerebro como el *hardware* que implementa un programa informático digital y considerar que la mente no manifiesta más libre albedrío que ese programa ejecutado en el *hardware*. Podríamos generar en la mente la ilusión de que tiene

libre albedrío si diseñáramos un programa con algunos elementos aleatorios o impredecibles, pero aun así el conjunto del sistema seguiría siendo determinista.

Hipótesis 2: el indeterminismo y el cerebro cuántico

La primera hipótesis es tranquilizante en este aspecto: el cerebro resulta ser una máquina como cualquier otra. Pero según la segunda hipótesis no está claro en modo alguno qué tipo de mecanismo deberá ser el cerebro a fin de que el sistema sea no determinista de la manera adecuada. Sin embargo, ¿cuál es exactamente la manera adecuada? Debemos suponer que la conciencia desempeña un papel causal en la determinación de nuestras decisiones y nuestras acciones libres, pero también que dicho papel causal no es determinista. Vale decir, no es una cuestión de condiciones suficientes. Ahora bien, la creación de la conciencia en cualquier instancia dada tiene que ver con las condiciones suficientes, por lo cual suponemos que los movimientos de izquierda a derecha de los procesos neurobiológicos a lo largo del tiempo no son en sí mismos causalmente suficientes. Esto es, ninguna etapa del proceso neurobiológico basta por sí misma para determinar la etapa siguiente a título de condiciones causalmente suficientes. Supongamos que la explicación de cada etapa por las precedentes depende de que todo el sistema sea consciente y tenga el tipo peculiar de conciencia que manifiesta un intervalo, es decir la conciencia voluntaria. ¿Cómo sería, empero, un sistema de esas características? Suponemos que en el nivel más básico el cerebro es no determinista, a saber, que el intervalo que es real en el nivel máximo desciende, por así decirlo, hasta el nivel de las neuronas y los procesos subneuronales. ¿Hay algo en la naturaleza que sugiera siquiera

la posibilidad de un sistema no determinista de ese tipo? La única parte de la naturaleza de la que en nuestros días, mientras escribo estas líneas, podemos afirmar con certeza que tiene un componente no determinista es la mecánica cuántica. Sin embargo, es un poco engañoso considerarla una parte, porque se trata del nivel más fundamental de la física, el nivel más básico de las partículas físicas. En el nivel cuántico, el estado del sistema en t_1 sólo es causalmente responsable de ese mismo estado en t_2 de una manera estadística y no determinista. Las predicciones hechas en el nivel cuántico son estadísticas porque hay un elemento aleatorio.

En el pasado siempre me pareció que la introducción de la mecánica cuántica en la discusión sobre el libre albedrío era totalmente irrelevante, por la siguiente razón: el libre albedrío no es equivalente al azar. La mecánica cuántica nos da azar, pero no libertad. Ese argumento me parecía convincente, pero ahora creo que cae en la falacia de la composición. (La falacia consistente en tomar propiedades de las partes de un sistema para atribuirlas a la totalidad de este.) Si suponemos que la creación de la conciencia por el cerebro es un resultado de procesos que, en algún nivel, son fenómenos cuánticos, y suponemos además que el proceso de deliberación consciente hereda la ausencia de suficiencia causal del nivel cuántico, no se sigue de ello, empero, que hereda el azar. Acaso sea posible que la función evolutiva de la conciencia consista al menos en parte en organizar el cerebro de tal manera que la toma de decisiones conscientes pueda producirse en ausencia de condiciones causalmente suficientes, aun cuando el efecto de la racionalidad consciente radica justamente en evitar las decisiones aleatorias. En pocas palabras, la aleatoriedad de los microprocesos que causan los

fenómenos conscientes en el macronivel no implica que estos sean aleatorios. Suponer lo contrario es incurrir en la falacia de la composición.

No obstante, decir que el libre albedrío es al menos posible si hay una explicación de la conciencia en términos de la mecánica cuántica no significa afirmar que ése es su funcionamiento real y ni siquiera su funcionamiento probable. Implica exclusivamente que, hasta donde sabemos, el único elemento no determinista establecido en la naturaleza es el nivel cuántico, y si debemos suponer que la conciencia es no determinista y el intervalo no sólo tiene realidad psicológica sino también neurobiológica, entonces, dada la situación actual de la física y la neurobiología, es preciso suponer que en la explicación de la conciencia hay un componente de la mecánica cuántica. No veo manera alguna de evitar esta conclusión.

Claro está, la hipótesis 2, según la cual la indeterminación aleatoria en el nivel cuántico conduce a una indeterminación de tipo no aleatorio en el nivel intencional consciente, parece muy improbable y muy poco convincente. Si se nos da a elegir entre la primera y la segunda hipótesis, pero también si se tiene en cuenta todo lo que sabemos de la naturaleza, la primera parece mucho más plausible.

Ocupémonos ahora, entonces, de examinar las ventajas y desventajas de ambas. La primera parece mucho más convincente y, en rigor, se ajusta a las descripciones del cerebro presentadas en los manuales clásicos de neurobiología. El cerebro es un órgano como cualquier otro. Está compuesto de células, y los procesos mediante los cuales estas se relacionan entre sí son tan deterministas como cualquier otro proceso celular, aun cuando, por supuesto, el cerebro tiene un tipo pe-

culiar de célula, la neurona, y en su interior hay relaciones peculiares entre ellas. Las neuronas se comunican a través de un notable proceso denominado potencial de acción, producido en las sinapsis. ¿Puede decirse algo contra la primera hipótesis? El único argumento que se me ocurre en su desmedro no es que se contrapone a nuestras experiencias de libertad (después de todo, tenemos toda clase de experiencias ilusorias), sino que muestra la experiencia del intervalo como si se tratara de una casualidad evolutiva, una especie de fenotipo evolutivo sin significado. La existencia del intervalo no es un rasgo fenotípico menor, como la existencia del apéndice. Que tengamos esas masivas experiencias de libertad y que estas carezcan de un valor biológico concreto parece un resultado absurdo desde el punto de vista de la evolución. El intervalo implica una gran inversión biológica de organismos como los seres humanos y los animales superiores. Una gran parte de la economía biológica del organismo está dedicada a la toma de decisiones racionales y conscientes. En el caso de los humanos esto tiene un aspecto tanto diacrónico como sincrónico de enorme magnitud. A lo largo de los años empleamos una vasta cantidad de tiempo, esfuerzo, dinero, etc., en prepararnos para tomar las mejores decisiones, y capacitamos a nuestros hijos con ese mismo fin. Pero si todos los detalles de nuestras presuntas decisiones libres ya están escritos en el libro de la historia en el momento del *big bang*, si todo lo que hacemos está enteramente determinado por fuerzas causales que actúan sobre nosotros, si la experiencia de la toma de decisiones libres y racionales es en todos los aspectos una ilusión, ¿por qué es una parte tan ubicua de nuestra historia vital biológica? ¿Y por qué es tan diferente de todo lo demás que conocemos

en la evolución? Este me parece el único argumento sólido que puedo presentar contra la primera hipótesis. Esta se contrapone a lo que sabemos de la evolución.

Es bastante fácil presentar argumentos contra la segunda. De hecho, esta hipótesis parece tan extraña a primera vista que resulta de inmediato poco convincente. Niega que el cerebro sea un órgano como cualquier otro y atribuye un papel especial a la toma de decisiones libres y conscientes. Ahora bien, hemos visto que no hay dualismo alguno en el hecho de que la conciencia pueda desempeñar un papel causal en la determinación de nuestro comportamiento. No estamos obligados a adoptar ni el dualismo ni el epifenomenalismo, pero de todas maneras, aun cuando evitemos ambos errores, seguimos frente a una descripción muy extraña de la conciencia. En la introducción de este libro dije que haría tanto hincapié en las zonas de la ignorancia humana como en las zonas de entendimiento. Este caso me parece un sólido ejemplo de ignorancia. En realidad no sabemos cómo existe el libre albedrío en el cerebro, si es que existe. No sabemos por qué o cómo la evolución nos ha dado la inconmovible convicción del libre albedrío. Y no sabemos, en síntesis, cómo puede llegar a funcionar. Pero sí sabemos que la convicción de nuestra libertad es inexorable. Si no la supusiéramos seríamos incapaces de actuar.

VI. Conclusión

El problema del libre albedrío nos va acompañar durante mucho tiempo. Los diversos esfuerzos para eludirlo, como el compatibilismo, no hacen sino permitirle reaparecer en otra forma. Aun después de haber resuelto las cuestiones más fundamentales abordadas

en este libro, interrogantes como cuál es la naturaleza de la mente, cómo se relaciona esta con el resto del mundo, cómo puede existir la causación mental y cómo puede nuestra mente tener intencionalidad, todavía sigue en pie la cuestión de si efectivamente tenemos libertad o no.

CAPÍTULO 9

EL INCONSCIENTE Y LA EXPLICACIÓN DEL COMPORTAMIENTO

Uno de mis principales objetivos en este libro es explicar cómo encajan los fenómenos mentales –conciencia, intencionalidad, causalidad y todos los otros rasgos de nuestra vida mental– en el resto del universo. Por ejemplo, ¿cómo existe la conciencia en un universo enteramente consistente de partículas físicas en campos de fuerza? ¿Cómo pueden los estados mentales funcionar de manera causal en ese universo? Hasta aquí, la mayor parte de nuestra investigación se ocupó de los fenómenos mentales conscientes. En este capítulo comenzaremos un serio examen de la naturaleza y el modo de existencia de los estados mentales *inconscientes*.

I. Cuatro tipos de inconsciente

Comencemos haciendo una pregunta ingenua: ¿los estados mentales inconscientes tienen existencia real? ¿Cómo puede haber un estado que es literalmente mental y al mismo tiempo inconsciente por completo? Los estados de esas características carecerían de cualitatividad y subjetividad y no formarían parte del campo unificado de la conciencia. Entonces, ¿en qué sentido, si lo hay, serían estados *mentales*? Y si cosas semejantes tienen existencia efectiva, ¿cómo pueden funcionar en términos causales como estados mentales y ser a la vez inconscientes? Nos hemos acostumbrado tanto a hablar del inconsciente, estamos tan cómodos con la idea de que además de los estados mentales conscientes hay estados mentales inconscientes, que hemos olvidado el

carácter enigmático que, en realidad, tiene la noción de inconsciente. Para Descartes, la pregunta: ¿existen los estados mentales inconscientes?, tiene una respuesta obvia. La idea de un estado mental inconsciente es una contradicción en sí misma. Descartes define la mente como *res cogitans* (cosa pensante) y "pensamiento" sólo es para él otro nombre de la conciencia. La idea de un estado mental inconsciente sería, por lo tanto, la idea de una conciencia inconsciente, una lisa y llana contradicción en los términos. Durante mucho tiempo, la idea cartesiana de la conexión necesaria entre lo mental y la conciencia gozó de suma influencia. El concepto y la importancia de los estados mentales inconscientes recién comenzaron a ser de aceptación general en el siglo pasado. Suele darse a Freud la mayor parte del crédito por esa aceptación, pero sus ideas tienen indudables precursores en Nietzsche y algunas figuras literarias, entre las cuales Dostoievski es quizá la de mayor trascendencia.

Entonces, ¿qué es exactamente un estado mental inconsciente? ¿Qué son, por ejemplo, una creencia o un deseo inconscientes? Creo que mucha gente, incluyendo a algunos autores sumamente sofisticados como el propio Freud, se forjó la siguiente imagen, bastante simplista. Un estado mental inconsciente es exactamente igual a un estado mental consciente menos la conciencia. El inconveniente de esta imagen es que cuesta mucho atribuirle algún sentido. Para verlo, el lector puede hacer esta prueba: piense de manera consciente "George Washington fue el primer presidente de Estados Unidos". Ahora, haga exactamente lo mismo, pero de manera inconsciente. Reste la conciencia. No tengo idea de cómo sería hacerlo o cuál es el presunto significado de la instrucción. No obstante, parece que no

podemos prescindir de la noción de inconsciente, por lo cual mejor será tratar de explicarla.

Mi estrategia en este capítulo, como en los anteriores, consistirá en comenzar con casos simples y no problemáticos para apoyar luego sobre ellos los casos más difíciles y desconcertantes. Empecemos con algunos casos no problemáticos de atribución de estados mentales a personas, en los que esa atribución no es en el acto un estado consciente. Para tomar un ejemplo obvio, de mí puede decirse sin lugar a dudas, aunque esté profundamente dormido, que creo que George Washington fue el primer presidente de Estados Unidos. Ahora bien, ¿qué hecho corresponde a esta aserción? ¿Qué hecho tocante a mí permite decir verazmente que tengo esa creencia hasta cuando no estoy consciente? Adviértase, por otra parte, que podemos atribuir la creencia de que George Washington fue el primer presidente de Estados Unidos incluso a una persona que está despierta por completo y piensa en algo totalmente distinto. Insistamos, entonces: ¿qué hecho corresponde a esas afirmaciones? Nótese que ninguna de ellas es una atribución enigmática y controvertida de inconsciencia. El propio Descartes habría aceptado la verdad de ambas. En los dos casos el hecho correspondiente a las afirmaciones es que en el hombre hay una estructura capaz de producir el estado en una forma consciente. Si cuando está despierto le preguntamos, por ejemplo, quién fue el primer presidente de Estados Unidos, es capaz de dar la respuesta correcta porque está en condiciones de producir el pensamiento consciente en cuestión. Debe advertirse que en este caso hemos identificado una estructura no en virtud de sus rasgos estructurales intrínsecos sino de lo que ella es capaz de causar. Este tipo de atribución es muy común en toda clase de casos no

problemáticos de la vida real. De una sustancia contenida en una botella decimos que es un limpiador, lejía o veneno sin examinar con mayor profundidad su estructura química. Simplemente la identificamos por lo que hace, no por la estructura que le permite hacerlo; y ahora sugiero que cuando decimos que el hombre tiene la creencia inconsciente de que George Washington fue el primer presidente de Estados Unidos, identificamos una estructura en él, no en virtud de sus rasgos neurobiológicos intrínsecos, sino a través de lo que ella hace, a través del estado consciente que es capaz de causar.

En estos casos hemos identificado un tipo de estado mental inconsciente, un tipo no problemático que Freud describió como "preconsciente".

Un segundo tipo de estado mental inconsciente es más problemático. A menudo sucede que un agente tiene estados mentales que actúan de manera causal en su comportamiento, pero él ignora por completo su funcionamiento e incluso puede negarlo sinceramente. Algunos de estos casos pertenecen al tipo que Freud describió como represión. En términos más generales, sin embargo, podemos caracterizarlos, también por medio del vocabulario freudiano, como inconsciente dinámico. Se trata de casos en los que el estado mental inconsciente, a pesar de serlo, funciona causalmente. Un ejemplo de estilo freudiano es el caso de Dora, que contrae una tos debido a su deseo sexual inconsciente por *Herr* K[1]. Los ejemplos freudianos suelen ser pro-

1 S. Freud, *Fragment of an Analysis of a Case of Hysteria*, en *Collected Papers*, vol. 3, Nueva York, Basic Books, 1959, pp. 13-146, sobre todo p. 49 y siguientes [traducción española: *Fragmento de análisis de un caso de histeria*, en *Obras completas*, vol. 7, Buenos Aires, Amorrortu editores, 1978].

blemáticos y gran parte de la obra clínica de Freud es a mi juicio científicamente inadecuada. Tomemos, no obstante, algunos casos en los que existen en realidad pocas dudas acerca de la exactitud científica de la descripción. En el capítulo anterior consideramos un ejemplo de hipnosis en el cual el agente actuaba claramente impulsado por un motivo desconocido para él y que presuntamente habría negado de pedírsele una explicación. En ese caso el hombre tenía el deseo de obedecer la siguiente orden: "Abra la ventana cuando escuche la palabra 'Alemania'", aun cuando desconocía que se le hubiese impartido dicha orden e ignoraba también todo deseo de cumplirla. De conformidad con Freud, designaremos los ejemplos de este segundo tipo como casos de estados mentales inconscientes reprimidos.

Un tercer tipo de estado mental también es objeto de un análisis muy frecuente en la literatura dedicada a la ciencia cognitiva. Se trata de casos en los cuales el agente no sólo no puede llevar el estado mental a la conciencia en los hechos, sino que ni siquiera podría hacerlo en principio, porque no es la clase de cosa susceptible de constituir el contenido de un estado intencional consciente. Así, por ejemplo, en la ciencia cognitiva suele decirse que un niño aprende un lenguaje a través de la aplicación "inconsciente" de muchas reglas de cómputo de una gramática universal, o es capaz de percepción visual gracias a que lleva a cabo operaciones de cálculo "inconscientes" relacionadas con el estímulo entrante a sus retinas. En ambos casos, tanto en la adquisición del lenguaje como en la formación de percepciones, las reglas de cómputo distan de

ser el tipo de cosas susceptibles de pensarse conscientemente. En última instancia, se reducen en su totalidad a secuencias masivas de ceros y unos, y cualesquiera sean las cosas que el niño puede hacer cuando piensa, es incapaz de pensar con ceros y unos; en rigor, estos sólo son una manera de hablar. Los ceros y unos existen en la mente del observador y constituyen un modo de descripción de lo que sucede de forma inconsciente en la mente del niño. Demos a estos casos, en los que el agente opera con reglas que no sólo son inconscientes de hecho, sino que jamás podrían ser conscientes, el nombre de "inconsciente profundo".

Además de estos tres tipos, hay una cuarta forma de fenómeno neurobiológico que no es consciente. En el cerebro suceden toda clase de cosas, muchas de las cuales tienen un papel crucial en el control de nuestra vida mental pero no son en modo alguno casos de fenómenos mentales. Así, por ejemplo, la secreción de serotonina en la hendidura sináptica no es, sin lugar a dudas, un fenómeno mental. La serotonina tiene una participación importante en varios tipos de fenómenos mentales, y algunas drogas de renombre, como el Prozac, se utilizan específicamente para inducir su secreción, pero su comportamiento no tiene como tal una realidad mental. Caractericemos como "no conscientes" este tipo de casos. Hay otros ejemplos de fenómenos no conscientes que son más problemáticos. Cuando estoy totalmente inconsciente, la médula sigue controlando mi respiración. Por eso no me muero en esa situación o cuando estoy profundamente dormido. Pero no hay realidad mental de los sucesos de la médula que me permiten respirar incluso cuando estoy inconsciente. No sigo inconscientemente la regla "siga respirando"; antes bien, la médula funciona de una manera

no mental, como lo hace el estómago cuando digiero comida.

Para resumir, entonces, hemos identificado cuatro tipos de fenómenos inconscientes: el preconsciente, el inconsciente reprimido, el inconsciente profundo y el no consciente. A mi entender, el primero y el cuarto no plantean problemas. ¿Qué pasa con el segundo y el tercero? En las siguientes secciones argumentaré que los casos de represión deben entenderse según el modelo del primero, el preconsciente; por su parte, los casos inconscientes profundos, del segundo tipo, se entienden de acuerdo con el modelo del cuarto, los casos no conscientes.

II. El principio de conexión

Me ocupo ahora de los casos de represión. Nuestra pregunta es esta: ¿cómo puede un estado mental reprimido existir y actuar como estado mental cuando es completamente inconsciente? Bien, ya vimos la respuesta en el caso del preconsciente. Adjudicar un estado mental a una persona en un momento en que el estado es inconsciente es atribuirle una estructura –cuyos detalles pueden ser totalmente desconocidos– que es capaz de producir ese estado en forma consciente. No es difícil, en realidad, decir que tal o cual persona dormida cree que George Washington fue el primer presidente, y tampoco lo es atribuir toda clase de creencias a un individuo consciente, aun cuando este no piense en ellas en el momento de la atribución. Ahora bien, me parece que este método funciona tambien para la segunda clase de casos, los de represión. Si digo que Sam actúa motivado por una hostilidad reprimida hacia su hermano o que Wolfgang actúa impulsado por

el deseo inconsciente de cumplir la orden que se le ha dado durante la hipnosis, en ambos casos les atribuyo una estructura neurobiológica capaz de causar un estado mental en forma consciente.

Pero con ello nos vemos ante lo que parece ser el problema más arduo. ¿Cómo pueden esos estados inconscientes, cuando lo son, causar un comportamiento humano concreto? ¿Cómo explicamos el "inconsciente dinámico"? A mi entender, cuando atribuimos esos estados mentales inconscientes a un agente, le atribuimos rasgos neurobiológicos capaces de causar conciencia. No sólo son capaces de causar estados conscientes sino un comportamiento consciente e incluso inconsciente. La cuestión, empero, es cómo puede el estado funcionar causalmente como estado mental en un momento en que sólo hay una estructura neurobiológica inconsciente. Como hemos hecho antes con otras cuestiones difíciles, la manera de responder a esta consiste en ocuparse ante todo de los casos simples y más evidentes.

Una vez me fracturé la muñeca. Durante el día la lesión me causaba bastante dolor, y este aumentaba si yo no tenía cuidado al mover el brazo. En el sueño noté algo interesante. Solía dormir completa y profundamente, de manera que no sentía dolor alguno; no obstante lo cual los movimientos de mi cuerpo durante la noche procuraban proteger la lesión. ¿Cómo describiríamos ese caso? ¿Debemos decir que al dormir yo tenía un dolor inconsciente y este me llevaba a comportarme de manera tal de no agravarlo? ¿O será preciso decir, al contrario, que mientras estaba profundamente dormido no tenía ningún dolor y el aparato neurobiológico subyacente capaz de causarlo en forma consciente actuaba causalmente en mí con el fin de impedir todo estímulo doloroso? Me parece que los hechos son los mismos en uno y otro caso. Por lo común no hablamos de dolores inconscientes, pero podríamos referirnos sin dificultades a ellos y casos como el descrito nos darían un motivo para hacerlo. Nótese que en este caso la neurobiología es capaz de causar el dolor en forma consciente, si bien durante un sueño profundo no siento conscientemente ningún dolor. Sin embargo, y aquí llegamos al punto crucial para esta parte del análisis, la neurobiología que es capaz de causar el dolor en forma consciente también es capaz de causar el comportamiento apropiado para evitarlo, incluso cuando no lo siento. Ahora bien, esto es justamente lo que necesito para describir los casos del inconsciente dinámico reprimido. Cuando este está en actividad, el agente no es consciente de ninguna motivación. No obstante, hay una estructura neurobiológica capaz a la vez de causar la aparición de la motivación como parte de los pensamientos conscientes del agente y de causar el comportamiento adecuado a ella. La única diferencia entre este caso y el ejemplo del dolor es que el agente quizá tenga razones adicionales para no querer reconocer la motivación. Pero –y esta es la respuesta que propongo a la cuestión– el modo de existencia, la ontología de la motivación inconsciente, cuando es inconsciente, es la de una estructura neurobiológica capaz de causarla en forma consciente, así como de causar el comportamiento apropiado para ella. Por eso, de paso, los freudianos ponían tanto afán en llevar lo inconsciente a la conciencia. Mientras lo inconsciente persista en ese estado, no está bajo nuestro control. No podemos reflexionar sobre él, ni juzgarlo, ni evaluarlo, ni someterlo a la racio-

nalidad, como sí podemos hacer de ordinario con las motivaciones existentes como parte de nuestros procesos de pensamiento racional consciente en el intervalo.

Hasta aquí, entonces, he señalado en este capítulo que hay casos absolutamente no problemáticos del inconsciente, los casos que denominamos preconscientes. Aun alguien como Descartes podría aceptarlos. Pero también he sostenido, de manera más polémica, que dichos casos proporcionan el modelo adecuado para considerar los ejemplos de represión, cuando el "inconsciente dinámico" está en funcionamiento. Sugiero que el mismo tipo de procesos neurobiológicos pueden causar también el comportamiento pertinente para ese estado consciente. De modo que hemos asimilado los primeros dos tipos de casos de inconsciencia a lo que ya sabemos sobre el cerebro y su funcionamiento, así como a lo que conocemos de nuestra vida mental consciente. En lo relacionado con la noción de inconsciente se disipan todos los misterios metafísicos, al menos para esta clase de casos.

Pero dediquémonos ahora a nuestro tercer tipo de casos, los correspondientes al inconsciente profundo. La tesis respectiva puede enunciarse con toda sencillez: esos casos no existen. No hay nada que pueda caracterizarse como un estado mental inconsciente profundo. Hay procesos neurobiológicos *no* conscientes que podemos describir *como si* fueran intencionales, y hay procesos neurobiológicos susceptibles de producir estados en forma consciente; pero cuando el estado mental no es siquiera el tipo de cosa que podría llegar a convertirse en contenido de un estado consciente, no se trata de un auténtico estado mental. Hemos analizado estos casos como si la neurobiología fuese intencional, como si fuera mental, como si siguiera reglas, pero no es así.

Mi tesis es que sólo entendemos un estado mental inconsciente como un estado que, aunque no consciente de inmediato, es capaz de llegar a serlo; y cuando lo atribuimos a un agente, describimos un mecanismo cerebral, no en términos de sus propiedades biológicas neurales, sino de su capacidad de causar estados y conductas conscientes. Doy a esta concepción el nombre de "principio de conexión", porque afirma que nuestra noción del inconsciente está lógicamente conectada con el concepto de conciencia. Un estado mental inconsciente debe ser el tipo de cosa que eventualmente podría ser un estado mental consciente[2].

¿Cuál es el argumento en apoyo de esta conclusión de apariencia pasmosa? En nuestra explicación de la intencionalidad (capítulo 6) vimos que todos los fenómenos intencionales tienen *formas aspectuales*. Pero en el caso del inconsciente profundo no hay nada de eso. No existe forma alguna de los estados intencionales que determine un contenido intencional en detrimento de otro. El argumento que planteo aquí es que debemos asimilar el tercer tipo de inconsciente, el inconsciente profundo, al cuarto, el no consciente, porque los casos de inconsciente profundo no tienen la característica esencial de los fenómenos intencionales, la forma aspectual del estado intencional que le permite actuar en la causalidad mental y por lo tanto justificar las formas mentalistas de explicación causal. No hay estados mentales inconscientes profundos. Antes bien, hay rasgos neurobiológicos que se comportan como si tuvieran intencionalidad.

¿Qué hay de malo en limitarse a decir que los procesos desplegados en el cerebro son estados intencio-

2 J. R. Searle, *The Rediscovery of the Mind*, *op. cit.*

nales inconscientes que aparecen en el acto como tales? ¿Por qué tenemos que pasar por un elaborado análisis disposicional consistente en señalar que la atribución de intencionalidad inconsciente es como la descripción de algo en cuanto veneno o lejía? La respuesta es que, como tal, la neurobiología no tiene una forma aspectual. Podremos entenderlo si consideramos algunos ejemplos. Imaginemos a un hombre que quiere tomar agua. Ahora bien, quizá sienta deseos de agua pero no de H_2O, simplemente porque no sabe que aquella es H_2O. No obstante, el comportamiento externo será exactamente el mismo en ambos casos: el del deseo de agua y el del deseo de H_2O. En uno y otro nuestro individuo procurará beber la misma clase de sustancia. Los deseos, empero, son diferentes. ¿Cómo debe aprehenderse la diferencia en el nivel de la neurofisiología? Descrita en términos de fuerza sináptica y potenciales de acción, la neurofisiología no sabe nada de formas aspectuales. Sin embargo, insistimos en decir que el hombre que tiene un deseo inconsciente de agua se encuentra en un estado intencional diferente del hombre que tiene un deseo inconsciente de H_2O, aun cuando la manifestación de ese deseo en la forma de un comportamiento sea exactamente la misma en los dos casos. La respuesta que propongo, y en rigor la única que me parece podría tener algún sentido, es que describimos la estructura neurobiológica desde el punto de vista de su capacidad de causar pensamientos y conductas conscientes. En el caso de la persona que no sabe que el agua es H_2O, la neurobiología correspondiente al deseo "quiero agua" es diferente de la neurobiología correspondiente al deseo "quiero H_2O". De todas maneras, en el nivel neurobiológico, estas diferentes formas aspectuales no existen como formas aspectuales sino,

por ejemplo, como diferencias de estructura neuronal. Por lo tanto, podemos asignar un sentido legítimo a la noción de inconsciente, con la condición de describirlo en términos de las capacidades causales del cerebro de generar conciencia.

Pero esto tiene una interesante consecuencia. Significa que no tenemos noción del inconsciente salvo en términos de lo consciente. Algo que no es siquiera el tipo de cosa que podría llevarse a la conciencia no puede ser un estado intencional, porque no puede tener forma aspectual. Por esa razón, no hay estados mentales inconscientes profundos. Hay estructuras neurobiológicas capaces de causar estados conscientes y comportamientos apropiados a esos estados mentales –y estos abarcan tanto los estados inconscientes reprimidos como los estados preconscientes– y hay estructuras neurobiológicas capaces de causar comportamientos que se manifiestan *como si* fueran intencionalmente motivados, pero en los cuales el tipo de motivación no puede ser un contenido intencional consciente y, por ende, no tiene realidad psicológica.

He dado un análisis disposicional de los estados mentales inconscientes. Un estado mental inconsciente, cuando lo es, consiste en la capacidad del cerebro de producirlo en una forma consciente, y de producir asimismo el comportamiento apropiado para él. Pero este resultado tiene una consecuencia inesperada para nuestro anterior análisis de la intencionalidad. He distinguido entre la red de estados intencionales y el contexto de capacidades que permiten el funcionamiento de estos. Sin embargo, ¿cuáles son los elementos de la red cuando son inconscientes? ¿Cuál es, por ejemplo, el estatus de mi creencia de que George Washington fue el primer presidente cuando estoy profundamente dor-

mido? Según el análisis disposicional que acabo de exponer, consiste en una capacidad cerebral. Pero si es así, el contexto también está compuesto por capacidades semejantes. Resulta entonces que la red de intencionalidad, cuando es inconsciente, es una subclase de las capacidades contextuales: la capacidad especial de producir ciertas formas de pensamientos y comportamientos conscientes.

III. Razones inconscientes para la acción

El tópico del inconsciente difiere de la mayoría de los demás temas ya discutidos en este libro en cuanto no se lo experimenta de inmediato; se trata, en cambio, de algo que hemos considerado necesario postular con algún otro fin. ¿Por qué es tan importante para nosotros? ¿Por qué nos interesa proponer una descripción del inconsciente, cuando por definición este ni siquiera puede experimentarse?

La respuesta es que el inconsciente ha llegado a ocupar un lugar de gran magnitud en nuestra explicación del comportamiento humano. Si lo postulamos es porque queremos explicar nuestro comportamiento. He escuchado a algunos filósofos afirmar que la razón por la cual decimos que la gente tiene creencias y deseos es que de esa manera podemos explicar su conducta. Para ser sincero, me parece que decir eso es tan poco inteligente como sostener que la razón por la cual decimos que la gente tiene pies es que de esa manera podemos explicar su comportamiento de marcha. No, si decimos que tienen pies es porque los tienen, y si decimos que tienen creencias y deseos también es porque los tienen. Pero la postulación del inconsciente es realmente parte de una necesidad explicativa. La razón por la cual decimos que las personas tienen motivaciones inconscientes radica en que no hemos encontrado otro modo de explicar algunas formas de su comportamiento. A diferencia de la "postulación" de los pies, o de las creencias y los deseos, la postulación de los estados mentales inconscientes *se hace* efectivamente con una finalidad externa: la explicación del comportamiento humano. Por eso tropezamos con un problema especial con respecto a la ontología del inconsciente, y por eso vale la pena hacer el esfuerzo de alcanzar una descripción de este que sea compatible con nuestra concepción global del mundo físico y el papel de lo mental en él.

Pero si necesitamos la noción de estados mentales inconscientes para explicar el comportamiento humano, nos hace falta una concepción previa de este y su explicación antes de saber cómo aplicar el concepto de inconsciente. He propuesto, al menos de manera preliminar, una descripción de la estructura de la acción humana en el capítulo 6, "La intencionalidad". Ese capítulo tiene ciertas implicaciones para la explicación de las acciones humanas, algunas de las cuales quiero presentar ahora.

El concepto clave para la explicación de una acción humana es el de razón. En nuestro análisis de la causalidad mental vimos que el contenido de la explicación debe concordar con el contenido presente en la mente del agente cuyo comportamiento se explica. Este es un punto de asombrosa importancia para disciplinas como la historia y las ciencias sociales, pero está disfrazado por la enorme complejidad de las explicaciones concretas. Decimos entonces, por ejemplo, que el alza de las tasas de interés norteamericanas causó un alza del valor del dólar. Y en la superficie el fenómeno parece muy simple: algo así como decir que el ascenso de la

temperatura causó un aumento de la presión. Pero en realidad, la explicación en términos de tasas de interés es inmensamente complicada. Para presentarla de manera exhaustiva, deberíamos explicar que la percepción de tasas de interés más elevadas en Estados Unidos impulsa a los inversores a desear invertir en valores norteamericanos con el fin de obtener mayores ganancias gracias a ellas, y que ese deseo, a su vez, genera la ambición de comprar más dólares para financiar las inversiones. Así pues, cuando digo que el contenido intencional de la explicación debe concordar con el contenido intencional de la mente de los agentes cuyo comportamiento se procura explicar, no pretendo señalar que haya una mera concordancia uno a uno en la superficie de la explicación.

¿Cuál es entonces la razón para una acción? Aunque la pregunta parece muy simple, la respuesta es de una enorme complejidad, y para exponerla con cierto detalle deberíamos ir más allá del alcance de este volumen. De hecho, he escrito un libro sobre ello (*Rationality in Action*, Cambridge (Mass.), MIT Press, 2000), de modo que el lector puede buscar en él los pormenores. Aquí sólo diré lo siguiente: si nos preguntamos cómo explicamos nuestro propio comportamiento, por ejemplo, por qué votamos al candidato que votamos en la última elección, comprobaremos que nuestras respuestas se incluyen en dos categorías. O bien aludimos a algún tipo de motivación, por ejemplo "quería impuestos más bajos", o mencionamos algún hecho que a nuestro entender está relacionado con esa motivación, por ejemplo, "creí que los republicanos bajarían los impuestos". Considerado en conjunto, este complejo forma lo que llamo una "razón total". Las razones siempre tienen una forma proposicional y para considerarlas como tales deben ser parte de una razón total. El punto clave para el examen del inconsciente es este. Hay algunas formas de comportamiento humano que sólo tienen sentido si postulamos una razón para la acción de la cual el agente mismo es inconsciente.

Una subcategoría especial de las razones para la acción son las reglas que gobiernan el comportamiento humano, y una forma especial de causalidad intencional se da en el comportamiento gobernado por reglas. El agente hace lo que hace, al menos en parte, porque sigue una regla. Pero ¿qué significa seguir una regla?

IV. Cumplimiento inconsciente de reglas

La capacidad explicativa de la postulación de los procesos mentales inconscientes depende en gran medida del supuesto de que esos procesos son casos de *cumplimiento inconsciente de reglas*. La idea es que nuestro comportamiento inteligente se explica a través de una multitud de procesos mentales inconscientes consistentes en el cumplimiento de reglas que ignoramos y no podríamos llegar a conocer. Pero si pretendemos entender la noción de cumplimiento inconsciente de reglas, es preciso comprender ante todo el concepto de cumplimiento de reglas, lo cual parecería requerir la comprensión de su cumplimiento consciente. ¿Qué hacemos exactamente cuando realizamos una acción como producto del cumplimiento de una regla? La respuesta a esta pregunta dista de ser obvia. Para explorarla, tendremos que especificar algunas de las características del cumplimiento de reglas. La primera distinción necesaria, y crucial para todo lo que sigue, es la existente entre el comportamiento *gobernado por reglas* y el comportamiento *descrito por reglas*. El com-

portamiento gobernado u orientado por reglas es aquel en el cual el agente que sigue la regla sufre la influencia causal de esta en su comportamiento. La regla actúa causalmente en la producción del comportamiento mismo constituido por su cumplimiento. Así, por ejemplo, si sigo la regla "maneje por el lado derecho de la ruta", su contenido debe funcionar causalmente y producir mi comportamiento. Esto no equivale a decir que el comportamiento está íntegramente determinado por la regla. Nadie sale a manejar un automóvil por el mero hecho de cumplirla, no obstante lo cual su contenido debe actuar causalmente; de no suceder así, significa que uno no la cumple. En este aspecto, el comportamiento en cumplimiento de reglas difiere del comportamiento descrito por ellas. Así, la pelota que se desliza por el plano inclinado puede ser descrita mediante las reglas de la mecánica newtoniana, pero de ello no se deduce que esté en algún sentido siguiéndolas. Su comportamiento en esa ocasión es descrito por reglas, pero no corresponde decir que se produce para cumplirlas.

¿Cuáles son, entonces, las características del comportamiento en cumplimiento de reglas? Enumeremos algunas de ellas.

1. Como acabamos de enunciar, el contenido de la regla debe actuar causalmente en la producción del comportamiento.
2. Debido a la característica 1, las reglas tienen las propiedades lógicas comunes a los estados intencionales volicionales y los actos de habla directivos. Por eso suele hacerse una analogía entre seguir una regla y obedecerla. Específicamente, las condiciones de satisfacción de esta tienen una

dirección de ajuste del mundo a la regla. El comportamiento debe cambiar para coincidir con el contenido de la regla. Esta también tiene la autorreferencialidad causal que, como vimos antes, es característica de las intenciones previas y las intenciones en la acción. La regla sólo es seguida si causa el comportamiento constituido por el hecho de seguirla.

3. De 1 y 2 se deduce que toda regla debe tener un contenido intencional que determina cierta forma aspectual. Podríamos tener entonces reglas de extensión equivalente cuyas condiciones de cumplimiento no fueran en modo alguno equivalentes. En mi automóvil, por ejemplo, la regla "maneje por el lado derecho de la ruta" daría el mismo resultado que "maneje de tal manera que el volante esté cerca de la línea divisoria de la ruta, y el asiento del pasajero, cerca del bordillo". Dada la estructura de los autos norteamericanos, esta regla producirá exactamente el mismo resultado que la regla inicial, pero una y otra, aunque equivalentes en extensión, no son iguales porque tienen diferentes formas aspectuales.
4. El cumplimiento de reglas suele ser voluntario. Para poder orientar el comportamiento, la regla debe permitir al agente seguirla voluntariamente. El intervalo, en resumen, está presente en el comportamiento gobernado por reglas. Por eso las "reglas" de acuerdo con las cuales digiero hidratos de carbono, por ejemplo, no son casos de cumplimiento de reglas sino de comportamiento descrito por ellas. Esto se debe a que no depende de mí. En suma, una de las características del cumpli-

miento de reglas es que estas pueden ser seguidas o rotas. Pero cuando no se las puede romper, tampoco se las puede seguir.

5. Como cualquier otro contenido intencional, las reglas siempre están sujetas a diferentes interpretaciones. Siempre es posible proponer otra interpretación para entenderlas. Así, por ejemplo, la mayoría de las reglas del comportamiento humano son lo que a veces se denomina reglas "cuando todo lo demás permanece constante" o *ceteris paribus*. Y esto se debe a que están sujetas a interpretaciones. De tal modo, yo respeto efectivamente la regla "maneje por el lado derecho de la ruta", pero no por eso interrumpo la marcha al enfrentarme con un obstáculo que bloquea ese lado del camino; me desvío para rodearlo por la izquierda. *Interpreto* la regla a fin de permitirme hacer cosas que no están especificadas en su contenido.
Esta característica del cumplimiento de reglas, el hecho de estar siempre sometidas a diferentes interpretaciones, ha dado pábulo a cierta forma de escepticismo. De acuerdo con una interpretación del famoso argumento de Wittgenstein sobre el lenguaje privado, este filósofo sostiene que cualquier comportamiento puede llegar a ser compatible con una regla con la condición de que nos tomemos la libertad de interpretar esta última[3]. Y

3 L. Wittgenstein, *Philosophical Investigations*, Nueva York, Macmillan, 1958 [traducción española: *Investigaciones filosóficas*, Barcelona, Crítica, 1988]; cf. S. Kripke, *Wittgenstein on Rules and Private Language*, Cambridge (Mass.), Harvard University Press, 1982 [traducción española: *Wittgenstein, reglas y lenguaje privado*, México, UNAM, 1989].

su respuesta a esto, según algunas interpretaciones, consiste en decir que nuestro cumplimiento de la regla es una práctica social, y la sociedad hace posible llegar a un acuerdo sobre lo que constituye cumplirla. Por esa razón, se atribuye a Wittgenstein haber mostrado que un "lenguaje privado" sería imposible, porque no habría control público de las interpretaciones de la regla.

6. El cumplimiento humano consciente de las reglas procede en tiempo real. Cuando respeto la consigna "maneje por el lado derecho de la ruta", la regla actúa causalmente en mi tiempo psicológico real para determinar condiciones de satisfacción. Mientras se trate de este sentido corriente del cumplimiento de reglas, será imposible que haya, por ejemplo, miles de reglas de cómputo que yo siga de manera más o menos instantánea, a semejanza de una computadora digital comercial. Su cumplimiento siempre exige cierta duración y procede en tiempo real.

Las anteriores son las características paradigmáticas del cumplimiento consciente de reglas. Pero cuando postulamos su cumplimiento inconsciente (y este tipo de postulaciones es muy común), ¿cuántas de esas características podemos mantener? Si hablamos literalmente del cumplimiento de reglas, esas son las características que debemos preservar. Si la referencia al cumplimiento inconsciente de reglas debe tomarse al pie de la letra, ese cumplimiento tiene que tener las siguientes características: la regla funciona causalmente con una dirección de ajuste del mundo a la regla y una dirección de causación de la regla al mundo. Las reglas

deben tener una forma aspectual, cumplirse de manera voluntaria y seguirse de modo tal que queden sujetas a diferentes interpretaciones; y es preciso cumplirlas en tiempo real. Algunas postulaciones de su cumplimiento inconsciente, como el correspondiente a la realización de los actos de habla, satisfacen esas condiciones. Pero no ocurre lo mismo con muchas otras postulaciones, como las descripciones de la ciencia cognitiva sobre la percepción visual y la adquisición del lenguaje.

V. Conclusión

La conclusión de este capítulo es un tanto deprimente. La noción de inconsciente es una de las concepciones más confusas e insensatas de la vida intelectual moderna. No obstante, parece que no podemos seguir adelante sin ella. En consecuencia, será preciso tratar de elaborar una noción coherente del inconsciente, que podamos ajustar a nuestros conocimientos sobre el resto de la realidad, incluido el funcionamiento del cerebro. El resultado es el principio de conexión. La mayoría de las personas que trabajan en este campo objetan mi definición de ese principio, pero no he visto que presentaran ninguna concepción alternativa coherente del inconsciente. Como conclusión, es lícito seguir utilizando esta noción, pero debemos reconocer que la usamos como una noción disposicional. Decir que un agente tiene tal o cual estado intencional inconsciente y que este interviene activamente en la producción de su comportamiento, significa decir que dicho individuo tiene un estado cerebral capaz de causar ese estado en forma consciente, aun cuando en alguna instancia específica pueda ser incapaz de hacerlo debido a un daño

cerebral, una represión, etc. No estoy del todo satisfecho con esta conclusión, pero no se me ocurre ninguna alternativa superior a ella.

CAPÍTULO 10

LA PERCEPCIÓN

Una de las principales funciones de la mente, tanto en nuestra vida cotidiana como durante la prolongada trayectoria evolutiva, consiste en relacionarnos con el resto del mundo, sobre todo por conducto de la percepción y la acción. Para expresar la cuestión en los términos más simples posibles, mediante la percepción recogemos información del mundo, luego la coordinamos consciente e inconscientemente y tomamos decisiones o forjamos intenciones, que resultan en acciones a través de las cuales afrontamos ese mismo mundo. En este capítulo consideraremos las relaciones entre la percepción y el mundo existente al margen de nuestras percepciones, lo que los filósofos dan en llamar, engañosamente, "mundo externo".

¿Por qué se supone que hay un problema allí? Si extiendo el brazo hacia adelante, veo mi mano frente a mi cara. Nada parecería más sencillo, ¿no es cierto? Hay una distinción tripartita entre yo mismo, la mano y la experiencia consciente y concreta de percepción por cuyo intermedio veo la mano. Podría contarse, desde luego, toda una compleja historia neurobiológica sobre el reflejo de la luz que, de la mano, va a dar en el sistema visual y desencadena una serie de procesos neuronales a cuyo término resulta la experiencia consciente de visión de la mano. Por otra parte, como constatamos en la discusión de la intencionalidad, hay algunas sutilezas filosóficas acerca de la forma de la autorreferencialidad causal implicada en las condiciones de satisfacción de la experiencia visual. Pero hasta aquí

esto no parece muy difícil. Sin embargo, debo decir al lector que en la historia de la filosofía pocas cuestiones causaron mayores trastornos que el problema de la percepción.

I. Argumentos en apoyo de la teoría de los datos de los sentidos

La concepción de la percepción que acabo de bosquejar es una forma de realismo perceptivo, a veces llamado "realismo directo" y en otras ocasiones "realismo ingenuo". La mayoría de los grandes filósofos en la historia del tema están convencidos de que es falso. Creen (y cuando hablo en plural me refiero a filósofos tan grandes como Descartes, Locke, Berkeley, Hume y Kant) que no vemos el mundo real. No vemos objetos y situaciones del mundo con existencia independiente. En realidad, todo lo que percibimos en forma directa –es decir, sin la mediación de ningún proceso inferencial– son nuestras propias experiencias internas. En el siglo pasado los filósofos solían expresar esta idea diciendo: "No percibimos objetos materiales, sólo percibimos datos de los sentidos". En la terminología inicial utilizada para denominar esos datos se cuentan palabras como "ideas" (Locke), "impresiones" (Hume) y "representaciones" (Kant). Pero si se preguntaba: "¿Cuál es el objeto directo de un verbo de percepción, tomado literal, estricta y filosóficamente?", la tradición respondía casi siempre que los objetos directos de los verbos de percepción no son expresiones designadoras de objetos materiales con existencia independiente, sino expresiones que aluden a nuestras propias experiencias internas, nuestros datos sensoriales.

El argumento de la ciencia

La descripción científica de la percepción muestra que las terminaciones de los nervios periféricos son estimuladas por objetos del mundo, y esa estimulación envía señales al sistema nervioso central y por último al cerebro; en este, el conjunto de los procesos neurobiológicos causa una experiencia perceptiva. Pero el único objeto real de nuestro conocimiento es esa experiencia en el cerebro. No hay manera de tener un acceso directo al mundo externo. Sólo podemos tenerlo al efecto que ese mundo tiene sobre nuestro sistema nervioso.

Este argumento parece presuponer que cuando describimos la estimulación de nuestras terminaciones nerviosas por los objetos del mundo hablamos de la percepción concreta del mundo real; de hecho, sin embargo, el argumento llega a la conclusión de que esa percepción es imposible. Hace más de sesenta años, Bertrand Russell enunció irónicamente esta aparente paradoja: "El realismo ingenuo", dijo, "conduce a la física, y la física, si es verdadera, muestra que el realismo ingenuo es falso. En consecuencia, el realismo ingenuo, si es verdadero, es falso; por lo tanto, es falso"[1].

A mi juicio, Russell señala que el realismo ingenuo parece de algún modo contraproducente. Si intentamos tomar con seriedad la idea de que estamos en contacto perceptivo directo con el mundo externo, y hacemos ciencia sobre esa base, la ciencia nos hará saber, como resultado, que no podemos estar en contacto perceptivo directo con el mundo externo.

---

1 B. Russell, *An Inquiry into Meaning and Truth*, Londres, Allen and Unwin, 1940, p. 15 [traducción española: *Investigación sobre el significado y la verdad*, Buenos Aires, Losada, 2003].

Creo que el argumento más susceptible de convencer a la mayor cantidad de gente en la historia de este tema es el de la ciencia. Pero en la historia de la filosofía el argumento que ha tenido más influencia entre los filósofos se denomina "argumento de la ilusión".

El argumento de la ilusión[2]

Si tratamos de tomarlo en cuenta con seriedad, el realismo ingenuo parece conducirnos a una suerte de inconsistencia y contradicción en los términos. Veamos por qué. Supongamos que tengo un cuchillo en la mano y lo veo. Pero Macbeth, en una situación mucho más dramática, también tuvo la experiencia de ver un cuchillo o, más específicamente, un puñal. Sin embargo, en ese momento tenía una alucinación. No veía un puñal real sino un puñal alucinado. En su caso, entonces, no podemos decir que viera un objeto material. Pero decididamente vio algo. Podríamos decir que vio la "apariencia de un puñal" o un "puñal alucinado". Ahora bien, y este es un paso crucial, si en el caso de Macbeth vamos a decir que sólo vio la apariencia de un puñal, deberíamos decir otro tanto en todos los casos, porque no hay diferencia cualitativa entre el carácter de la experiencia en los episodios verídicos y en los episodios alucinatorios. Por eso Macbeth se engañó: no había diferencias entre su experiencia en esos momentos y la de ver realmente un puñal. Pero si decimos que en todos los casos sólo vemos una apariencia y no el objeto

2 En A. J. Ayer, *The Foundations of Empirical Knowledge*, Londres, Macmillan, 1953, se encontrará una exposición de diferentes versiones del argumento de la ilusión.

mismo, con seguridad deberemos encontrar un nombre para esas apariencias. Llamémoslas "datos de los sentidos". Conclusión: nunca vemos objetos materiales sino únicamente datos de los sentidos. Y entonces surge esta pregunta: ¿cuál es la relación entre los datos de los sentidos que vemos y los objetos materiales que al parecer no vemos?

Esta forma de argumento ha circulado en una amplia variedad de ejemplos. Aquí tenemos otro. Cuando pongo un dedo frente a mi cara y concentro la vista en la pared del otro extremo del cuarto, se produce un fenómeno conocido como doble visión. Veo duplicado el dedo. Ahora bien, aunque lo veo doble, no veo dos dedos. Sólo hay uno. Es obvio, sin embargo, que veo dos unidades de algo. ¿Dos unidades de qué? Démosles el nombre de apariencias de un dedo; y en verdad veo dos apariencias de un dedo. Pero si así son las cosas –otro paso crucial–, no hay diferencias cualitativas entre ver las apariencias de un dedo y ver el dedo real. Puedo probármelo si modifico el foco para que ambas apariencias se unan. Donde antes veía dos apariencias, ahora veo una sola. En consecuencia, si pretendemos decir que en el caso de la doble visión sólo vemos apariencias y no objetos materiales, deberíamos decir lo mismo en todos los casos. Busquemos un nombre para esas apariencias: llamémoslas "datos de los sentidos".

Ahora, un tercer argumento. Si pongo una varilla recta dentro de un vaso de agua, la varilla, debido a las propiedades refractivas de la luz, parece torcerse. Sin embargo, no está realmente torcida, sólo parece estarlo. Sea como fuere, cuando la miro veo directamente algo torcido. ¿Qué es? Veo directamente la apariencia de una varilla, y esta exhibe en efecto un aspecto torcido. Pero la varilla misma no está torcida; la que lo está es la apa-

riencia. Insistamos, no obstante: lo que veo sin mediaciones está torcido, y por lo tanto se trata de la apariencia y no de la varilla. A esta altura el lector ya debe saber cuál es el próximo paso: si en este caso voy a decir que no veo la varilla sino la apariencia, debería decir lo mismo en todos los casos, porque no hay diferencia cualitativa entre ellos. Necesitamos una expresión para describir esas apariencias. ¿Adivinan cuál es? Las llamaremos "datos de los sentidos". Conclusión: nunca veo objetos materiales, sólo veo datos de los sentidos.

Podría seguir todo el día con estos ejemplos, pero daré sólo un par más para que el lector saboree en todos sus matices el estilo argumentativo. Supongamos que me levanto de la silla y camino alrededor de la mesa manteniendo los ojos fijos en ella. Mientras camino, algo cambia; más aún, cambia algo que yo percibo directamente. No se trata de la mesa, que permanece absolutamente inalterada mientras la rodeo. ¿Dónde están los cambios, entonces? Naturalmente, en la apariencia de la mesa. Esta me presenta una apariencia diferente desde distintos puntos de vista. Ahora bien, como lo que veo está cambiando y no se trata de la mesa, y lo que veo es la apariencia, me parece que sólo veo apariencias y no la mesa. Por otra parte, como no hay distinción cualitativa entre esta experiencia y cualquier otra, al parecer estoy obligado a concluir que nunca veo otra cosa que apariencias. Necesitamos una expresión técnica para denominarlas. Las llamaremos "datos de los sentidos".

A continuación otro ejemplo, también célebre. Saco una moneda del bolsillo y la sostengo en alto. Cuando la miro de frente parece redonda. Pero si la doy vuelta levemente para que me ofrezca un ángulo, deja de parecer redonda; ahora se muestra elíptica. Ahora

bien, estamos seguros de una cosa: la moneda misma no es elíptica. No ha cambiado su forma por haberla inclinado un poco. También estamos seguros, sin embargo, de que percibo directamente algo elíptico. Es innegable que aquí, en mi campo visual, hay algo elíptico; lo veo en forma directa. Pero al parecer, entonces, no estoy viendo la moneda, porque esta es redonda. Lo que veo directamente, lo que veo sin ningún proceso inferencial, es la apariencia elíptica de la moneda. Y si voy a decir que en este caso sólo veo apariencias, debería repetir lo mismo en todos los casos, porque cuando pongo la moneda bien derecha, para que me presente una apariencia redonda y no elíptica, no hay un cambio cualitativo. La conclusión es evidente: deberíamos decir en todos los casos que veo apariencias, no objetos materiales, y esas apariencias pueden denominarse "datos de los sentidos".

Casi todos los filósofos célebres de los últimos trescientos cincuenta años, así como la mayor parte de los filósofos respetables hasta mediados del siglo XX, aceptaron de una u otra manera la teoría de los datos de los sentidos. Hume, a decir verdad, creía que el realismo ingenuo era tan notoriamente falso que apenas se molestó en refutarlo. En un momento dice que si uno siente la tentación de adoptar ese realismo, para refutarlo le basta con apretarse un globo ocular. Cuando lo hacemos vemos todo doble; según Hume, el realista ingenuo tendría que concluir que el número de objetos contenidos en el universo simplemente se ha duplicado. Pero como es obvio que no ha sido así, nuestro filósofo cree lícito deducir que no vemos los objetos materiales[3].

3 D. Hume, *A Treatise of Human Nature*, *op. cit.*, pp. 210-211.

El argumento de la ilusión tiene una estructura lógica común a todos estos ejemplos. Podemos describirla del siguiente modo:

1. Los realistas ingenuos suponen que, al menos en el caso típico, vemos objetos materiales, y los vemos como realmente son.
2. Pero hay muchos casos, como aun el realista ingenuo admitiría, en los que no vemos objetos materiales (por ejemplo, en los episodios alucinatorios) o no los vemos como realmente son (así ocurre, por ejemplo, en los casos de la varilla torcida y la moneda elíptica).
3. Sin embargo, aun en esos casos vemos algo y lo vemos como realmente es. En las situaciones en que no hay absolutamente ningún objeto material, como el ejemplo del puñal de Macbeth, este veía algo. Había algo directamente presente en su campo visual. Y en los casos en que hay un objeto material pero no lo vemos como es en realidad, tal cual sucede en los ejemplos de la moneda elíptica y la varilla torcida, vemos algo elíptico y vemos algo torcido. Tanto la entidad elíptica como la entidad torcida tienen una presencia directa en nuestro campo visual.
4. En esos casos vemos directamente apariencias, etc. (datos de los sentidos), y no objetos materiales.
5. Desde un punto de vista cualitativo, estos casos no difieren del caso estándar; por ende, si con respecto a ellos decimos que vemos datos de los sentidos y no objetos materiales, deberíamos decir lo mismo en todos los casos.

II. Consecuencias de la teoría de los datos de los sentidos

La doctrina del realismo directo sostiene que, al menos de manera habitual, percibimos en forma directa los objetos y situaciones del mundo. Negamos esta doctrina cuando decimos que nunca percibimos esos objetos y situaciones, sino nuestras propias experiencias, los datos de nuestros sentidos. Pero una vez asumida esa actitud nos enfrentamos a una cuestión muy seria: ¿cuál es la relación entre los datos de los sentidos que sí percibimos y los objetos que al parecer no percibimos? Aunque en la historia de la filosofía han sido muchas las respuestas dadas a esta pregunta, a mi juicio se las puede reducir a dos familias fundamentales. Una, la que despierta una atracción más inmediata, consiste en decir que no percibimos los objetos mismos sino *representaciones* de los objetos. El dato sensorial que sí percibimos es una especie de imagen del objeto, de modo que podemos tomar conocimiento de este si inferimos su presencia y sus rasgos de las características de los datos de los sentidos. El objeto concreto en el mundo real se asemeja a estos últimos al menos en ciertos aspectos. Algunos filósofos –el más importante tal vez sea Locke– trazaron una distinción entre los rasgos de los datos de los sentidos que tienen elementos semejantes correspondientes en el mundo real y los que no los tienen. Los rasgos del mundo real que se asemejan efectivamente a los datos sensoriales recibieron el nombre de "cualidades primarias" y entre ellos se incluyeron la forma, el tamaño, el número, el movimiento y la solidez. (La lista de Locke es "solidez, extensión, figura, movimiento o reposo y número"[4].) Pero hay

4 J. Locke, *An Essay Concerning Human Understanding*, edición es-

otros datos de los sentidos para los cuales no hay un rasgo semejante correspondiente en el mundo real. De manera engañosa, Locke llamó a estos rasgos de los objetos "cualidades secundarias". Y digo "de manera engañosa" porque, estrictamente hablando, esas cualidades de los objetos no existen. Antes bien, como señala el propio Locke, las cualidades secundarias sólo son las capacidades de las cualidades primarias de causar en nosotros ciertas experiencias. Esas cualidades secundarias son el color, el olor, el sabor y el sonido. Nuestras experiencias de ambos tipos de cualidades son provocadas por rasgos reales del objeto; pero el objeto mismo no tiene los rasgos correspondientes a nuestras experiencias de las cualidades secundarias.

Esta doctrina se denomina teoría representativa de la percepción y fue elaborada con bastante detalle, sobre todo por Locke. Según sus términos, pasamos nuestra vida consciente como si estuviéramos dentro de un cine. Podemos ver imágenes del mundo real en la pantalla, pero nunca podemos ir más allá para ver el propio mundo real, porque el cine está en su totalidad dentro de nuestra mente. Todo lo que vemos son más imágenes y más representaciones. Tanto Berkeley como Hume atacaron, creo que con mucha eficacia, la teoría representativa. El ataque puede adoptar varias formas, pero el argumento básico, el argumento para el cual no parece haber una respuesta, es el siguiente: si decimos que nuestros datos sensoriales se asemejan a los objetos y por eso los representan a la manera como la escena de una película representa la escena real, tropezamos

tablecida por A. S. Pringle-Pattison, Oxford, Clarendon Press, 1924, p. 67 [traducción española: *Ensayo sobre el entendimiento humano*, México, Fondo de Cultura Económica, 1992].

con el inconveniente de no contar con un significado claro de la noción de "semejanza" y, por consiguiente, tampoco de "representación". ¿Cómo podemos decir que los datos de los sentidos que vemos se asemejan al objeto que no vemos, si este último es, por hipótesis, totalmente invisible? Es como si yo afirmara que en mi garaje tengo dos autos exactamente iguales, pero uno de ellos es invisible sin remedio. No tiene ningún sentido decir que hay una relación de semejanza perceptiva entre algo que tiene rasgos perceptivos y algo que no los tiene.

Contra lo que podría haberse supuesto, cuando Berkeley conoció esta objeción no volvió al realismo ingenuo ni dijo que debía haber cometido un error al pasar de la teoría de la percepción fundada en esa doctrina a la teoría de los datos de los sentidos. Antes bien, señaló que sólo existían las mentes y las ideas. El mundo real consiste enteramente de datos de los sentidos. No hay objetos materiales por añadidura a nuestras experiencias reales y posibles. Aunque de una manera más compleja, Hume llegó a una conclusión similar. Esta concepción tiene varios nombres, pero quizás el más común sea fenomenalismo. Los objetos materiales consisten en colecciones de datos sensoriales; no hay objetos materiales por encima de los fenómenos mentales o sumados a ellos.

El fenomenalismo pretendía ser una tesis lógica, por lo cual la manera más clara de enunciarlo es presentarlo como una tesis lógica sobre el lenguaje. En vez de decir que los objetos consisten de datos de los sentidos, afirmación que nos hace parecer en discrepancia con la idea de que están compuestos de moléculas, deberíamos decir, en realidad, que las proposiciones sobre los objetos e incluso las proposiciones empíricas

en general pueden traducirse sin pérdida de significado en proposiciones sobre los datos sensoriales. El mismo impulso de verificación que llevó al conductismo en la filosofía de la mente condujo al fenomenalismo en la filosofía de la percepción. Así como la única prueba que tenemos de la existencia de otras mentes es el comportamiento, la única prueba de los objetos materiales es al parecer la aportada por los datos sensoriales. En consecuencia, una concepción verdaderamente científica de la mente debe ser conductista; de manera análoga, una concepción verdaderamente científica del mundo material debe ser fenomenalista.

III. Refutación de la teoría de los datos de los sentidos

A mi entender, toda esta manera de concebir la percepción es desesperadamente errónea. Como dije antes, creo que es la más desastrosa teoría de la historia de la filosofía en los últimos cuatro siglos. ¿Por qué? Porque hace imposible dar una descripción veraz de la relación de los seres humanos y otros animales con el mundo real. Conduce de manera casi inevitable de Descartes y Locke a Berkeley y Hume, y de estos a Kant. Y luego las cosas se ponen verdaderamente feas cuando la tradición llega a Hegel y el idealismo absoluto. La mera idea de volver a atacarla me deprime enormemente, pero no habré cumplido la tarea que prometí al lector en este libro si no intento responderla punto por punto. A eso vamos, entonces.

Los argumentos en respaldo de la tesis de los datos de los sentidos son falaces sin excepción. Consideremoslos en orden.

El argumento de la ciencia

La ciencia no refuta el realismo ingenuo. Considerar nuestra capacidad de dar una descripción causal de la manera de ver el mundo real y deducir de ella que no vemos ese mismo mundo es caer en una célebre falsedad, la llamada falacia genética. Esta consiste en suponer que una descripción causal que explica la génesis de una creencia, su modo de adquisición, muestra con ello que la creencia es falsa.

La falacia genética suele referirse a creencias, pero su forma puede generalizarse. La idea es esta: si podemos mostrar que las causas de una creencia u otro contenido intencional son insuficientes para probar su verdad, de alguna manera refutamos dicha creencia u otro estado intencional.

En mi infancia intelectual, las formas más comunes de la falacia genética se encontraban en el freudismo y el marxismo. ¿El lector duda de la verdad del marxismo? Esa actitud sólo muestra que sus orígenes burgueses lo inducen a error. ¿Duda de la verdad de las enseñanzas de Freud? Su vacilación prueba únicamente que es víctima de su propia represión. En nuestros días la falacia genética no tiene mucha difusión, excepto en los posmodernistas. Yo solía preguntarme por qué era tan habitual en el posmodernismo hasta que leí un trabajo que explica por qué los posmodernistas no tienen realmente a su disposición otra forma de argumentación[5].

Sea como fuere, la forma de la falacia genética en la teoría de la percepción es la siguiente. Podemos

5 M. Bauerlein, *Literary Criticism: An Autopsy*, Filadelfia, University of Pennsylvania Press, 1997.

mostrar que cuando creemos vernos la mano frente a la cara, lo que sucede es, en realidad, que la luz reflejada de la primera causa en nosotros una experiencia visual, y suponemos que se trata de la experiencia visual de la mano. Como es posible explicar por qué creemos verla, es posible mostrar que en realidad no hemos visto una mano frente a nuestra cara sino su mera experiencia visual, efecto de los procesos neurobiológicos.

Así enunciada, espero que la falacia resulte evidente. La descripción causal del modo como llego a ver la mano frente a la cara no muestra que realmente no la veo.

El argumento de la ilusión

Es más complicado dar una réplica al argumento de la ilusión. Tomaré prestadas tanto las ideas como las técnicas de mi maestro en filosofía, J. L. Austin, a fin de refutar este argumento[6].

Adviértase que en todos los argumentos presentados hasta aquí, la estrategia lingüística consiste en conseguir un sustantivo que será el objeto directo de los verbos de percepción, pero que no designa un objeto material. Así, en el caso del puñal de Macbeth, se nos dijo que no veíamos un arma real sino un arma alucinada. Pero la dificultad de esta explicación es que en el sentido de "ver", yo realmente veo un cuchillo en mi mano; en el caso de la alucinación no veo nada. Expresiones como "puñal alucinado" no pueden designar una especie de puñal. Para decirlo en pocas palabras, cuando Macbeth tuvo una alucinación no vio nada. O, al menos, no vio nada perteneciente al rubro "puñales". Sin duda se vio las manos. Entonces, del hecho de que Macbeth tuviera una alucinación fenomenológicamente indistinguible de una experiencia real no se sigue que viera un tipo especial de objeto o entidad que es común a las experiencias verídicas e ilusorias.

6 J. L. Austin, *Sense and Sensibilia*, Oxford, Oxford University Press, 1962 [traducción española: *Sentido y percepción*, Madrid, Tecnos, 1981].

Objeciones similares pueden hacerse con respecto a los casos de doble visión. Nunca deberíamos aceptar la cuestión de manera acrítica. La pregunta era esta: cuando me veo el dedo duplicado, ¿veo dos unidades de qué? La respuesta es: cuando me veo el dedo duplicado, no veo dos unidades de nada; veo un dedo y lo veo doble.

Tanto en el ejemplo del dedo doble como en el de la varilla torcida, se introduce la noción de apariencia para proporcionar un objeto directo a los verbos de percepción. La idea es que no vemos el objeto mismo sino su apariencia. Pero si lo pensamos bien, hay algo autocontradictorio en la idea de que podríamos ver la apariencia de un objeto y no ver el objeto. Ver la apariencia de un objeto es simplemente ver su aspecto. Y no hay modo de ver el aspecto de algo sin ver ese algo. La consideración de algunos ejemplos aclarará por completo esta idea. Supongamos que pregunto: "¿Viste qué buen aspecto tenía Sally en la fiesta?" No tiene sentido que mi interlocutor me conteste: "Sí, vi que tenía buen aspecto pero por desdicha no pude verla a ella. Sólo pude ver su apariencia".

Apliquemos estas consideraciones al ejemplo de la mesa. Me levanto y camino alrededor de ella. Su apa-

riencia cambia, porque la veo desde diferentes perspectivas, pero ella misma no cambia; por lo tanto, parece que veo la apariencia y no la mesa. Espero que la falacia de este argumento resulte obvia. La mesa, desde luego, parece diferente desde distintos puntos de vista. Pero los cambios en mis experiencias visuales, provocados por el hecho de cambiar de posición y por lo tanto de perspectiva, no muestran que no pueda ver la mesa sino algo que, por decirlo de algún modo, se interpone entre ella y yo, su apariencia. Al contrario, toda la discusión presupone que siempre veo realmente la mesa, pues no habría manera de que esta siguiera mostrándome diferentes apariencias desde distintos puntos de vista si yo no la viera efectivamente.

El paso en falso crucial de la estructura argumental que he sintetizado es el tercero: en todos los casos percibimos algo y lo percibimos tal como realmente es. Esto no es cierto. En los casos de alucinación no percibimos nada, y en los otros –la varilla torcida, la moneda elíptica, etc.– percibimos el objeto, pero en condiciones que pueden ser más o menos engañosas. Del hecho de que la varilla parezca (un poco) torcida no se deduce que veamos una entidad torcida, el aspecto. No, realmente vemos una varilla, un objeto material con existencia independiente, que en esas condiciones parece torcido.

Es asombroso que estos argumentos hayan tenido tan grande influencia en la historia de la filosofía. En mi opinión, no resisten un mínimo escrutinio y dejo al lector, como ejercicio práctico, la tarea de ver cómo podríamos utilizar estas lecciones para mostrar la falacia en el caso de la moneda elíptica.

IV. Un argumento trascendental en apoyo del realismo directo

Alguien podría decir que la refutación de los argumentos contra el realismo ingenuo no basta para mostrar que este es verdadero. La objeción es atinada. Necesitamos algún argumento que demuestre que, al menos en ciertas oportunidades, percibimos efectivamente objetos materiales y situaciones del mundo. ¿Cuál podría ser?

El problema que enfrentamos aquí es una variante del escepticismo tradicional. El argumento del escéptico es siempre el mismo: podríamos contar con todas las pruebas con que contamos y, en rigor, con toda la evidencia posible, y pese a ello persistir en el error. Se nos insta a probar, por ejemplo, que realmente vemos la mesa frente a nosotros y no tenemos una mera alucinación, un sueño, somos víctimas de un genio maligno, etc. No hay manera de dar al escéptico una respuesta directa sobre mi presente experiencia visual de la mesa. El quid de su posición es que yo podría tener exactamente esa misma experiencia y de todas maneras estar equivocado. Y si puedo estar equivocado en este caso, ¿por qué no en todos?

No me parece filosóficamente astuto tratar de proponer una respuesta directa a este argumento. No me creo capaz de demostrar al escéptico que en este momento estoy viendo realmente la mesa y no alucinando, soñando, etc. Sí puedo mostrar, en cambio, que cierto estilo de discurso, el discurso que el escéptico suele adoptar, presupone la verdad de alguna versión del realismo directo. (Me gusta pensar que mi versión es "ingenua", pero no importa que sea ingenua o sofisticada.) El realismo en cuestión debe contener la idea

de que por lo menos en algunas ocasiones tenemos acceso perceptivo a los fenómenos públicamente observables. De ordinario, estos se conciben como "objetos materiales", pero esa designación, insistamos, no es crucial. Lo crucial es que diferentes personas puedan, al menos en ciertas oportunidades, percibir los mismos fenómenos públicamente observables: sillas, mesas, árboles, montañas, nubes, etc. El argumento que estoy por presentar es un argumento "trascendental" en uno de los muchos sentidos kantianos del término. En un argumento trascendental así entendido, suponemos que cierta proposición *p* es verdadera y luego mostramos que una de sus condiciones de posibilidad es que otra proposición *q* también lo sea. En este caso suponemos que hay un discurso inteligible públicamente compartido por distintos hablantes/oyentes. Suponemos que las personas se comunican efectivamente entre sí mediante un lenguaje público sobre objetos y situaciones públicas del mundo. Luego mostramos que alguna forma de realismo directo es una condición de posibilidad de esa comunicación. La clave del argumento reside en ver que la hipótesis de los datos sensoriales ha reducido, sin revelarlo en forma explícita, el mundo *públicamente* disponible de objetos materiales a un mundo *privado* de datos de los sentidos. Sólo yo puedo experimentar mis datos de los sentidos. Sólo usted puede experimentar los suyos. Pero si es así, ¿cómo podemos siquiera hablar del mismo objeto en un lenguaje público? ¿Cómo podemos, en síntesis, llegar a comunicarnos unos con otros sobre otros objetos públicos? Si los objetos materiales son reducibles a datos sensoriales, y los únicos datos sensoriales a los que tengo acceso son los míos propios, nunca podré

comunicarme con usted en lo concerniente a un objeto material público.

A continuación, los pasos del argumento:

1. Suponemos que, al menos algunas veces, logramos comunicarnos con otros seres humanos.
2. La comunicación en cuestión asume la forma de significados públicamente accesibles en un lenguaje público. En términos más específicos, cuando digo cosas como esta: "Esta mesa es de madera", supongo que usted entenderá las palabras del mismo modo que yo. Si no fuera así, no conseguiríamos comunicarnos.
3. Sin embargo, a fin de lograr comunicarnos en un lenguaje público, debemos suponer objetos de referencia comunes y al alcance de todos. Así, por ejemplo, cuando utilizo la expresión "esta mesa", tengo que suponer que usted la entiende tal como yo pretendo. Tengo que suponer que ambos nos referimos a la misma mesa, y cuando usted entiende mi enunciado de "esta mesa", considera que se refiere al mismo objeto al cual usted hace referencia en este contexto al pronunciar la frase "esta mesa".
4. Esto implica que usted y yo compartimos un acceso perceptivo al mismo objeto. Lo cual no es sino otra manera de decir que yo debo presuponer que ambos vemos o percibimos de algún otro modo el mismo objeto público. Un lenguaje público presupone un mundo público. Pero la disponibilidad pública de ese mundo es justamente el realismo directo que intento defender aquí. El inconveniente de la hipótesis de los datos de los

sentidos, como del fenomenalismo en general, es que ignora la privacidad de dichos datos. Una vez planteada la tesis de que no vemos los objetos públicamente disponibles sino los datos sensoriales, el solipsismo parece estar a la vuelta de la esquina. Si sólo puedo hablar de manera significativa de objetos que en principio están a mi alcance desde un punto de vista epistémico, y los únicos objetos en esa situación son los datos privados de los sentidos, es imposible que logre comunicarme en un lenguaje público, porque no tengo forma de compartir el mismo objeto de referencia con otros hablantes. A eso aludía cuando decía que un lenguaje público presupone un mundo público. Pero el supuesto de ese mundo público es precisamente el realismo ingenuo que he propiciado. No demostramos la verdad del realismo ingenuo; sólo probamos la ininteligibilidad de su rechazo en un lenguaje público.

CAPÍTULO 11

EL YO

En la célebre máxima de Descartes, "pienso, luego existo", ¿a qué se refiere la primera persona del singular? Para el filósofo, no se refiere ciertamente a mi cuerpo sino a mi mente, la sustancia mental que constituye mi yo esencial. Tenemos ahora una buena razón para suponer que el dualismo cartesiano no es una descripción filosóficamente aceptable de la naturaleza de la mente. Pero quienes rechazan el dualismo aún deben hacer frente a una cuestión seria: ¿qué es exactamente el yo? ¿Qué hecho correspondiente a mí me hace ser yo? Muchos filósofos contemporáneos, yo mismo entre ellos hasta hace bastante poco, creen que Hume dijo más o menos la última palabra sobre el tema. Además de la secuencia de experiencias y el cuerpo en el cual estas ocurren, no hay nada que pueda llamarse yo. Cuando dirijo la atención hacia mi interior y trato de descubrir alguna entidad que constituya lo esencial de mi persona, dice Hume, todo lo que descubro son experiencias particulares; no hay yo alguno junto a ellas.

El tema del yo plantea varias cuestiones más o menos independientes entre las cuales distinguiré, a los fines perseguidos en este capítulo, tres familias diferentes.

I. Tres problemas del yo

1. ¿Cuáles son los criterios de la identidad personal?

Un persistente interrogante tradicional en la filosofía ha sido el siguiente: ¿qué hecho hace que una persona sea la misma a través de los diversos cambios que sobrelleva en el curso de la vida? En mi caso, por ejemplo, he pasado por una cantidad bastante grande de cambios en las últimas décadas. Mi cuerpo tiene un aspecto un tanto diferente, he aprendido algunas cosas nuevas y olvidado algunas cosas viejas, mis aptitudes y gustos han experimentado diversas modificaciones, pero de todos modos es innegable que a través de todos esos cambios sigo siendo exactamente la misma persona. Soy idéntico a la persona que llevó mi nombre y vivió en mi casa décadas atrás. Pero ¿qué hace que la secuencia de sucesos y cambios que acabo de mencionar corresponda a la vida de una y la misma persona?

2. ¿Cuál es exactamente el sujeto de nuestra atribución de propiedades psicológicas?

Además de la secuencia de sucesos psicológicos que constituyen la percepción, la acción, la reflexión, etc., y el cuerpo en el cual esos sucesos se desarrollan, ¿debemos postular algo más?

No he formulado esta pregunta con demasiada precisión, pero intentaré hacerlo más adelante. Por el momento, mi intención es plantear una cuestión general: por añadidura a mi secuencia de pensamientos y sentimientos reales y el cuerpo en el cual estos ocurren, ¿es necesario postular una cosa, una entidad, un "yo" ["*I*"] que sea el sujeto de todos esos sucesos? Supongamos que todos podemos concordar, como he dado por sentado a lo largo de este libro, en que estoy constituido al menos en parte por un cuerpo físico, y que este contiene una secuencia de fenómenos mentales: estados conscientes y procesos cerebrales inconscientes capaces de producir estados conscientes. La pregunta es: ¿debemos postular algo más? Y si es así, ¿de qué se trata? Hasta donde yo sé, la mayoría de los filósofos contemporáneos siguen a Hume en la idea de que no tenemos que postular nada más; por mi parte, aunque con renuencia, me he visto obligado a reconocer que sí debemos hacerlo, y explicaré por qué en el curso de este capítulo.

3. ¿Qué es exactamente lo que hace de mí la persona que soy?

En la vida contemporánea suele considerarse que esta cuestión tiene que ver con fuerzas sociales, psicológicas, culturales y biológicas que modelan mi personalidad específica y hacen de mí la clase de persona que soy. En el habla popular hay, en expresiones como "política de la identidad" o "identidad cultural", un uso de la noción de "identidad" concerniente a las fuentes, tanto culturales como biológicas, que dan forma a la personalidad de cada uno. Creo que este sentido del concepto de identidad personal difiere del atribuido a la expresión en las preguntas 1 y 2. En este último caso el concepto está más vinculado con el carácter y la personalidad que con el problema metafísico de la existencia y la identidad de un yo a través del tiempo.

Este capítulo se ocupará de la familia de cuestio-

nes relacionadas con las preguntas 1 y 2. Veremos que nos plantean suficientes dificultades sin necesidad de abordar las cuestiones de la personalidad.

II. ¿Por qué hay un problema especial con respecto a la identidad personal?

Las cuestiones sobre la identidad son tan antiguas como la filosofía, pero parece haber un problema especial en lo relativo a la identidad de las personas. El más famoso enigma sobre la identidad en la historia del tema es probablemente el ejemplo de la "nave de Teseo". Durante un tiempo, una nave de madera es objeto de una reconstrucción completa y gradual. El barco sigue navegando, tiene una tripulación que lo hace surcar el Mediterráneo, pero poco a poco las planchas que lo conforman son reemplazadas una a una hasta que no queda nada de la construcción original. Ahora bien, ¿sigue siendo la misma nave? Bien, a mi juicio la mayoría estimaría que sí, que la continuidad espacial y temporal del funcionamiento es suficiente para garantizar su identidad como nave, porque el concepto de nave es, después de todo, una noción funcional. Supongamos ahora, sin embargo, que alguien recoge los maderos desechados y los utiliza para construir un barco que contiene todas las partes de la nave originalmente botada y sólo ellas, de manera que cada plancha del segundo barco es idéntica a una plancha del primero. ¿Cuál es la nave con la que partimos? ¿La que muestra continuidad de función o la que tiene continuidad de partes? El error en estos debates, como ocurre tantas veces en filosofía, es suponer que con respecto a la identidad debe haber alguna verdad adicional de los hechos, más allá de todos los datos que acabo de mencionar. A mi entender no existe ninguna otra verdad. Depende de nosotros decir cuál es la nave original. El asunto podría tener alguna importancia, por ejemplo, para decidir quién es el dueño de qué barco. ¿Quién es responsable de pagar los impuestos? ¿Cuál de las naves tiene derecho de muelle? Pero, más allá de los hechos que he enumerado, no queda ninguna otra cuestión fáctica con respecto a cuál de los barcos es idéntico al original.

Algunas de las cuestiones sobre la identidad personal son similares al caso de la nave de Teseo, pero cuando se trata de aquella sentimos que hay un problema especial, ausente en los ejemplos tradicionales. Solemos creer que cada uno de nosotros se presenta a sí mismo de una manera especial y que esas experiencias de primera persona son esenciales para nuestra identidad, mientras que los fenómenos de tercera persona son más o menos incidentales. Todos creemos entender, por ejemplo, qué significaría decir que una mañana podríamos despertar y descubrirnos en un cuerpo diferente. Como Gregor Samsa en el relato de Franz Kafka, nuestra apariencia física externa habría cambiado por completo, pero de algún modo sabríamos, aun cuando nadie más estuviera convencido de ello, que somos la misma persona que antes ocupaba otro cuerpo. Para hacer más concreto este ejemplo, supongamos que el transplante de cerebro se convierte en una posibilidad real y que el mío es transplantado en el cuerpo de Jones, y viceversa. Desde mi punto de vista me parece innegable que una vez realizada la intervención voy a pensar que soy exactamente la misma persona que antes, pero mi cerebro (y yo, por lo tanto) ocupará entonces un cuerpo diferente. Quizá me costará convencer de esto a otra gente, pero, al menos desde el punto de vista

de primera persona, sentimos sin lugar a dudas que yo me veré como el mismo individuo que antaño ocupaba un cuerpo distinto y ahora habita en el cuerpo de Jones.

Un caso más desconcertante: imaginemos que todas mis capacidades mentales se realizan de igual manera en ambos lados del cerebro. Imaginemos a continuación un caso de bisección cerebral y el transplante de cada uno de los hemisferios en un cuerpo diferente. Supondremos que el cuerpo original se deja a un lado y ahora las dos mitades de mi cerebro están implantadas en otros dos cuerpos. ¿Cuál de los personajes resultantes, si puedo describirlos así, corresponde a mí? Este caso me parece similar al ejemplo de la nave de Teseo, por cuanto no hay en la cuestión más hechos que los ya mencionados. Esto es, me parece que tenemos iguales razones para decir que soy el número uno o el número dos; o más probablemente digamos que ahora hay dos personas, cuando antes había una sola. Este caso es como los ejemplos de fisión, cuando una ameba se divide en dos. No obstante, desde el punto de vista de primera persona, aun en esta situación uno siente que debe haber una verdad de los hechos. Si ahora soy uno de los frutos de la fisión, es probable que diga: "Sigo siendo yo, el mismo individuo único que siempre fui. No me importa lo que digan los demás". El problema es que mi gemelo tendrá exactamente la misma convicción con la misma justificación, y los dos no podemos tener razón.

Una característica típica de nuestros conceptos es que su aplicación al mundo real presupone cierto tipo de regularidades. Esto es tan válido para los conceptos de barco, casa, árbol, automóvil o perro como para conceptos tan raros como el de identidad personal. Por lo común podemos recurrir a este último concepto por-

que los criterios de primera y tercera personas tienden a reunirse. No se distancian de manera radical. Pero es fácil imaginar mundos de ciencia ficción en los cuales lo hagan. Supongamos que la fusión y la fisión se tornan habituales; es decir, supongamos que fuera muy común la reunión repentina en un solo cuerpo de varias personas que caminan por la calle. O bien, para tomar el caso de la fisión, imaginemos que una sola persona pudiera ramificarse en cinco individuos idénticos como resultado de la fisión de su cuerpo original. Si tales casos llegaran a ser corrientes, tendríamos serios problemas con nuestra noción de identidad personal. Creo muy probable que ya no fuera válida.

III. Los criterios de la identidad personal

Si observamos concretamente los criterios utilizados por la gente en el habla cotidiana para decidir qué persona es hoy idéntica a qué persona del pasado, comprobamos la existencia de por lo menos cuatro condiciones que constituyen nuestra noción de identidad personal. Dos de ellas corresponden al punto de vista de tercera persona, una procede de la perspectiva de primera persona y la cuarta es mixta. Revisémoslas.

1. Continuidad espacio-temporal del cuerpo

Mi cuerpo es continuo en el espacio y el tiempo con el de una criatura nacida varias décadas atrás. Más que en cualquier otra cosa, el público se apoya en esa continuidad espacio-temporal para considerarme la misma persona. Adviértase que la continuidad espacio-temporal de mi cuerpo no implica la misma continuidad de las micropartes que lo componen. En el nivel

molecular, mis partes corporales sufren un proceso constante de reemplazo. Las moléculas que componen mi cuerpo son hoy totalmente diferentes de las presentes en el inicio de mi vida, pero, de todos modos, sí, sigue siendo el mismo cuerpo, sobre todo debido a su continuidad espacio-temporal con el cuerpo original del lactante.

2. Continuidad temporal relativa de la estructura

A pesar de que mi estructura cambia a través de las décadas –crezco y envejezco–, soy de todas maneras un ser humano reconocible. Si, como Gregor Samsa, despertara una mañana metamorfoseado en el cuerpo de un gran insecto, o me transformara de improviso en un elefante o una jirafa, no parece evidente que las otras personas estuvieran dispuestas a decir que sigo siendo John R. Searle. Por lo tanto, además de la mera permanencia en bruto de un continuo a través del espacio y el tiempo, al parecer también necesitamos reconocer ciertos tipos de regularidades estructurales en los cambios sufridos por ese objeto espacio-temporal.

Si la identidad personal plantea un problema especial es porque estas dos condiciones no parecen suficientes para mi punto de vista de primera persona. Aun cuando otros se nieguen a reconocer a mi persona en cierto objeto, confío en mi capacidad de saber, desde mi punto de vista interno de primera persona, quién soy, aunque me encuentre en el cuerpo de un elefante o una jirafa e incluso si me reduzco al tamaño de un pulgar; sea como fuere, podré autoidentificarme. Pero ¿a qué deben equivaler esos criterios?

El siguiente criterio es de primera persona.

3. Memoria

Desde mi perspectiva interna existe al parecer una secuencia continua de estados conscientes unidos por mi capacidad de recordar, en cualquier momento dado, experiencias conscientes ocurridas en el pasado. Muchos filósofos, y sobre todo Locke, consideraron que ése era el elemento esencial de la identidad personal. El motivo por el cual lo necesitamos por añadidura a esta última es que parece fácil imaginar casos en los que yo despertara en un cuerpo diferente, pero desde mi punto de vista seguiría siendo sin lugar a dudas el mismo. Aún tendría mis experiencias como parte de la secuencia. Esta incluye experiencias de recuerdo de estados conscientes pasados. Locke, al encontrar en ella la característica esencial de la identidad personal, la llamó "conciencia", pero la interpretación más difundida es que se refería a la memoria. Hobbes y Hume se creyeron en condiciones de refutar esa tesis señalando que las relaciones de la memoria eran intransitivas. Esto es, el viejo general podría recordar acontecimientos ocurridos cuando era un joven teniente y el joven teniente podría recordar sucesos de su infancia, pero el viejo general quizás hubiera olvidado la niñez. En este aspecto, Hobbes y Hume tenían seguramente razón, pero el hecho de que olvidemos cosas no parece representar una refutación de la idea de que desde el punto de vista de la primera persona, la secuencia de mis estados conscientes, unidos por la memoria, es esencial para discernir mi existencia como la de un individuo específico.

4. Continuidad de la personalidad

Este criterio tal vez sea menos importante que los otros tres, no obstante lo cual hay cierta continuidad relativa de mi personalidad y mis disposiciones. Si mañana a la mañana, al despertarme, me sintiera y me comportara exactamente como la princesa Diana poco antes de su muerte, cabría preguntarse si soy "realmente la misma persona". También podemos tomar un caso real, el famoso ejemplo de Phineas Gage, que sufrió un daño cerebral mientras trabajaba en un equipo de tendido de líneas ferroviarias y una barra de acero le atravesó el cráneo. Milagrosamente, Gage sobrevivió, pero su personalidad se trastocó por completo. Así como antes había sido una persona entusiasta y agradable, luego del accidente comenzó a mostrarse como un hombre vil, receloso, vicioso y desagradable. En cierto sentido, podríamos considerar que Gage era "otra persona". Adviértase, sin embargo, que al describir estos casos seguimos usando el mismo nombre propio que antes. A efectos prácticos, es innegable que continuamos hablando de Phineas Gage. En lo concerniente a asuntos cotidianos como determinar quién debe su impuesto a la renta o es el dueño de su casa, no juzgamos esencial la impresión de que se trata de otra persona. No obstante, sus amigos y su familia podrían sentir que "no es el mismo".

Tal como se señaló antes, la operatividad de un concepto depende de una diversidad de criterios que le otorgan validez, y el supuesto tácito antecedente es que todos ellos actúan juntos. Así sucede, en efecto, en los casos con que estamos familiarizados en la vida normal. De todas maneras, se plantean algunos enigmas.

IV. Identidad y memoria

He dicho que la memoria cumple un papel esencial en nuestra concepción de primera persona de la identidad personal. A continuación veremos por qué. Tengo hoy recuerdos conscientes de anteriores experiencias conscientes de mi vida, así como la capacidad de evocar un número muy grande de recuerdos similares de otras experiencias pasadas. La sensación de que soy exactamente el mismo individuo a lo largo del tiempo, desde mi punto de vista de primera persona, se debe en gran parte a mi aptitud de producir recuerdos conscientes de sucesos conscientes anteriores de mi vida.

Creo que a eso se refería Locke cuando dijo que la conciencia desempeña un papel esencial en nuestra concepción de la identidad personal, pero con prescindencia de que se tratara de eso o no, la continuidad de la memoria es por lo menos un aspecto importante de la idea que nos formamos de esa identidad. Leibniz planteó un argumento similar: imagina que llegas a ser emperador de China, pero has perdido toda huella de la memoria de tu pasado. No hay diferencia, dice Leibniz, entre imaginar esto e imaginar que dejas de existir y nace un nuevo emperador.

Hay una objeción tradicional al planteamiento de Locke, que mucha gente estima decisiva y que ahora quiero exponer y responder. Dice así: el planteamiento es circular. Sólo podemos decir verdaderamente que un agente es capaz de recordar sucesos de su vida anterior si presuponemos que es idéntico a la persona que vivió dichos sucesos. Pero, por lo tanto, no podemos explicar la identidad personal en términos de memoria, porque la memoria en cuestión presupone la identidad misma

que intentamos explicar. Podemos expresar esta objeción de manera más formal.

Una persona P_2 en el tiempo T_2 es idéntica a una persona anterior P_1 en el tiempo T_1 si y sólo si P_2 recuerda en T_2 sucesos ocurridos a P_1 en T_1, donde los hechos en cuestión son experiencias conscientes y la experiencia misma de recordar también lo es.

La afirmación sobre la circularidad de esta idea se justifica del siguiente modo: a fin de que P_2 recuerde realmente en T_2 un suceso ocurrido a P_1 en T_1, en contraste con el mero hecho de pensar que lo recuerda, P_2 debe ser idéntico a P_1. Pero si esto es cierto, no podemos utilizar la memoria para justificar la afirmación o el criterio de identidad, porque requerimos esta última como condición necesaria de la validez de aquella.

Podemos ilustrar estas observaciones con algunos ejemplos. Supóngase que ahora digo, sin faltar a la verdad, recordar haber escrito la *Crítica de la razón pura*. Esto no establece ni tiende a respaldar de manera alguna la idea de que soy idéntico a Immanuel Kant, porque sabemos que yo no podría haber escrito esa obra por no ser idéntico a él, que sí la escribió. Pero exactamente por el mismo motivo, si ahora digo que recuerdo haber escrito *Speech Acts*, esto no sirve de por sí para establecer que soy idéntico a John Searle, autor de ese libro, porque antes de poder saber que acierto al recordar haber escrito *Speech Acts* deberíamos saber que soy John Searle. Los dos casos son paralelos en todos los aspectos. ¿Este argumento es decisivo contra la teoría de que la memoria es una parte esencial de la identidad personal? A mi entender, la respuesta variará según cuál sea la pregunta que, a nuestro juicio, la teoría trata de responder. Si consideramos que esa pregunta es: ¿cuáles son los criterios de la identidad personal tales

que, de ser satisfechos, la persona P_2 en T_2 será idéntica a la persona P_1 en un tiempo anterior T_1?, el criterio no se cumple. Cualquiera sea la cantidad de recuerdos putativos de Kant que yo tenga, no por eso soy Kant. Sin embargo, hay una pregunta diferente que a mi parecer es respondida por la teoría, y se trata de la pregunta de primera persona: ¿qué hay en mí, en mis experiencias personales, que me lleva a sentirme yo mismo como una entidad continua a través del tiempo, agregada a la continuidad de mi cuerpo? Y con respecto a esta pregunta, me parece que la continuidad de mis experiencias de memoria es una parte esencial de la percepción de mí mismo como un yo continuo. Alguien que no fuera yo podría tener experiencias personales idénticas en su tipo que le dieran un sentido de sí mismo idéntico en su tipo al mío. De todas maneras, no seríamos idénticos, no obstante lo cual cada uno de nosotros se siente como un yo continuo.

V. Un argumento a favor de la existencia de un yo no humeano

Todas estas discusiones dejan en pie la cuestión de si necesitamos o no el concepto de un yo por añadidura a la noción de disposiciones y estados psicológicos. Creo que la mayoría de los filósofos coinciden con Hume en sus críticas a Locke y Descartes, en el sentido de que no hay un yo o una identidad personal más allá de la secuencia de nuestras experiencias reales. El escepticismo de Hume con respecto al yo es similar a su actitud escéptica acerca de la conexión necesaria y la causación. Nuestro filósofo observa a su alrededor para ver si puede descubrir alguna impresión unificadora de todas sus percepciones; no es una sorpresa que

no la encuentre. Cuando vuelco mi atención hacia adentro, nos dice, encuentro experiencias específicas. Descubro este o aquel deseo de tomar agua, un leve dolor de cabeza o la sensación de opresión de los pies dentro de los zapatos, pero ninguna experiencia del yo se suma a esas experiencias particulares. Por consiguiente, cualquier identidad que yo pueda atribuirme debe ser un resultado de la secuencia de experiencias específicas. Es una ilusión, dice Hume, suponer que por encima de estas hay algo que constituye mi yo. Como en el caso de la conexión necesaria, las palabras de Hume parecen dar a entender un lamentable fracaso de nuestra parte por no poder descubrir la experiencia del yo, así como no logramos encontrar la experiencia de aquella conexión. Sin embargo, como en esa otra cuestión, el filósofo plantea un argumento lógico y no un argumento psicológico sobre la ausencia de un tipo determinado de experiencia. El argumento es este: nada puede mostrarse como una experiencia del yo, porque cualquiera que tuviéramos, aun la que durara toda una vida, sería simplemente una experiencia más. Supongamos que yo tuviese una mancha amarilla constante en mi campo visual que me acompañara sin falta durante toda mi vida consciente. ¿Sería eso un yo? No, sólo sería una mancha amarilla. Nada podría satisfacer las condiciones necesarias para que algo fuera una experiencia del yo, esto es, una experiencia que unificara todas las demás. Creo que, en el nivel al cual se dirigen, los argumentos de Hume son muy convincentes, y me parece que muchos filósofos, tal vez la mayoría, coincide conmigo en cuanto al vigor que manifiestan.

Pero he llegado a regañadientes a la conclusión de que Hume dejó algo al margen; y esto nos lleva a nuestra segunda serie de preguntas: ¿necesitamos postular

algo además de nuestro cuerpo y la secuencia de nuestras experiencias? Me he visto en la necesidad de concluir que sí, debemos postular sin duda alguna un yo por añadidura a la secuencia de experiencias, y ahora presentaré un argumento en apoyo de ese postulado.

Volvamos a nuestro supuesto original de que yo consisto en un cuerpo y una secuencia de experiencias. Esta secuencia incluirá cosas como el sabor del café, la visión del color rojo, el panorama de la bahía de San Francisco desde mi ventana, etc. ¿Queda algo afuera? Creo que sí. Debemos notar en primer lugar algo que ya señalé. No tenemos experiencias desordenadas; antes bien, todas las experiencias que tengo en un instante cualquiera se viven como parte de un solo campo consciente unificado. Por lo demás, su poseedor experimenta la continuación de ese campo consciente a través del tiempo como una continuación de su propia conciencia. Vale decir, no siento que mi conciencia de hace cinco minutos y ni siquiera de hace cinco años esté desconectada de mi conciencia actual; vivo en cambio la experiencia de una conciencia continua interrumpida por fases de sueño. (Los debates filosóficos no toman en cuenta lo suficiente el fascinante hecho de que uno tenga una sensación del paso del tiempo incluso durante el sueño, al menos en este aspecto: al despertar, sentimos que ha pasado más o menos tiempo desde que nos dormimos. Al parecer, no sucede lo mismo con las personas que han quedado inconscientes a causa de un golpe o han recibido anestesia general.)

Los argumentos que me convencieron de que es preciso postular por lo menos una noción formal del yo (más adelante diré a qué me refiero al hablar de "formal") tienen que ver con las nociones de racionalidad, libre elección, toma de decisiones y razones para la

acción. En el capítulo 7 señalamos que las explicaciones intencionalistas de la toma de decisiones y la actuación racional humana tienen una forma lógica peculiar, diferente de la forma clásica de las explicaciones causales. El contraste es, por ejemplo, el existente entre estas dos frases:

1. Puse una x en la boleta electoral porque quería votar por Bush.
2. Me dio dolor de estómago porque quería votar por Bush.

Ahora supondremos, en beneficio del argumento, que ambas frases son ciertas y proporcionan explicaciones adecuadas. Pese a ello, su forma lógica es muy diferente. De acuerdo con una interpretación convencional, la número 2 enuncia condiciones causalmente suficientes. En ese contexto, mi deseo de votar por Bush fue suficiente para provocarme un dolor de estómago. Pero la frase número 1, también según una interpretación convencional, no propone condiciones causalmente suficientes. Sí, marqué la boleta con una X por esa razón, pero bien podría no haberlo hecho. Después de todo, podría haber decidido no votar por Bush, irme del recinto o hacer muchas otras cosas. Sin embargo, ahora nos encontramos en apariencia frente a un enigma. ¿Cómo puede ser adecuada la explicación de mi comportamiento en términos de razones, si no presenta condiciones causalmente suficientes? Sin ellas, no explica por qué hice lo que hice y no otra serie de cosas que perfectamente podría haber hecho, en igualdad de todas las demás condiciones. Al parecer, si la explicación no enuncia condiciones causalmente suficientes, no explica de manera adecuada el fenómeno

que pretendía dilucidar. Pero la respuesta decisiva a esta objeción es que, desde mi punto de vista, la explicación es absolutamente adecuada. Lo que explico es mi comportamiento, y puedo apelar a mis razones para explicar por qué hice lo que hice, sin comprometerme en modo alguno con la idea de que esas razones enuncian condiciones causalmente suficientes. En rigor, tal vez esté muy al tanto de que no lo hacen.

¿Cómo debemos interpretar entonces los enunciados de la forma 1 y, a decir verdad, cualquier otro que proponga una explicación de mi comportamiento voluntario libre a través de mis razones para actuar? La respuesta, creo, es que, además del "haz de percepciones", tal como las describió Hume, debemos suponer que ciertas coacciones formales se ejercen sobre la entidad que toma las decisiones y lleva a cabo las acciones. Tenemos que postular un yo o agente racional capaz de actuar libremente y hacerse responsable de las acciones. El complejo de las nociones de acción libre, explicación, responsabilidad y razón nos da la motivación para postular algo por añadidura a la secuencia de experiencias y el cuerpo en el cual estas ocurren. Para ser más exacto, a fin de explicar las acciones racionales libres, debemos suponer la existencia de una entidad X tal que de ella pueda decirse que es consciente (con todo lo que la conciencia implica), persiste a través del tiempo, formula razones para la acción y reflexiona sobre ellas bajo las coacciones de la racionalidad, es capaz de decidir, iniciar y llevar a cabo acciones bajo un supuesto de libertad y (ya implícito en lo que he dicho) es responsable de al menos algunos de sus actos.

Hume creía tener una objeción decisiva contra cualquier postulación semejante. No tengo experiencia alguna de ese yo, ese X. Si oriento introspectivamente

mi atención y examino todas las experiencias que tengo en este momento, no daría el nombre de "yo" a ninguna de ellas. Siento la tela de la camisa en la espalda, el regusto del café en la boca y una leve resaca por lo que bebí anoche y capto la vista de los árboles a través de la ventana, pero nada de esto es un yo ni podría considerarse como tal. Entonces, ¿en qué consiste ese yo? Creo que Hume tiene toda la razón: no hay experiencia de esa entidad, pero esto no significa que no debamos postularla o proponer algún principio formal; ahora examinaré con mayor profundidad qué tipo de razones nos obligan a ello y qué tipo de entidad podría ser el yo en cuestión.

Una manera de pensar estas cuestiones es concebirlas como problemas de ingeniería. Si estuviéramos diseñando un robot consciente y quisiéramos que reprodujera toda la gama de capacidades racionales humanas, es decir que fuese capaz de reflexionar sobre las razones para la acción, tomar decisiones y actuar bajo el supuesto de su propia libertad, ¿qué elementos tendríamos que incorporarle?

El primer requisito evidente de cualquier robot de esas características es que debería ser consciente. Por otra parte, la forma de su conciencia tendría que ser cognitiva, en el sentido de que debería tomar los estímulos perceptivos, procesar conscientemente la información recibida de la percepción y razonar sobre esa base en procura de llevar a cabo una acción.

Su segunda característica sería la capacidad de poner en marcha una acción, una capacidad a veces llamada "agencia". Se trata de una aptitud adicional a las percepciones conscientes, presente en los seres humanos y muchos animales. Es un rasgo de ciertos tipos de conciencia, pero no de todos. El paso crucial es, a mi

juicio, el tercero. El agente racional consciente que hemos creado debe ser capaz de embarcarse en algo que en inglés denominamos *acting on reasons* [actuar sobre la base de razones]. Ahora bien, esto es importante porque la noción de actuar sobre la base de una razón difiere de la idea de hacer que algo nos suceda causalmente. Ése era el sentido de la ilustración que presenté antes sobre la diferencia entre la afirmación de que me dio dolor de estómago porque quería votar por Bush y la de que realicé una acción libre, actué sobre la base de mi deseo de votar por Bush. La noción de "actuar sobre la base de" presupone el intervalo de libre albedrío descrito anteriormente. Hasta aquí, entonces, en nuestro robot hemos puesto conciencia, junto con experiencias perceptivas conscientes y otros estados intencionales, la capacidad de reflexionar sobre esos estados y la agencia racional, que es la capacidad peculiar de emprender acciones bajo un supuesto de libertad. Pero si hemos hecho todo eso, ya tenemos un yo. El yo que describo es un concepto puramente formal; no implica ningún tipo específico de razón o de percepción. Se trata, antes bien, de una noción formal que incluye la capacidad de organizar su intencionalidad bajo coacciones de racionalidad, de tal manera que sea posible realizar acciones voluntarias intencionales, cuyas razones no son causalmente suficientes para determinarlas.

¿Por qué esa noción del yo es "formal" y no "sustantiva"? Para responder a esta pregunta, me gustaría recurrir a una analogía entre el yo y otra noción formal. A fin de entender mis percepciones visuales, debo concebir que ocurren *desde un punto de vista*, pero por sí mismo este no es algo que veo o percibo de alguna otra manera. Es un requisito puramente formal necesario

para hacer inteligible el carácter de mis experiencias. El punto de vista no tiene rasgos sustantivos al margen de esta restricción formal, a saber, que debe ser el punto desde el cual se producen mis experiencias. Ahora bien, la noción de yo que postulo es, de manera similar, puramente formal, pero más compleja. Debe ser una entidad tal que, en su singularidad, tenga conciencia, percepción, racionalidad, la capacidad de lanzarse a la acción y la de organizar percepciones y razones a fin de llevar a cabo acciones voluntarias bajo un supuesto de libertad. Si tenemos todo eso, tenemos un yo.

Ahora podemos explicar muchas otras características, dos de las cuales tienen un papel central en nuestro concepto de yo humano. Una es la responsabilidad. Cuando me embarco en acciones comprometo mi responsabilidad, y de ese modo cuestiones como el merecimiento, la culpa, la recompensa, la justicia, el elogio y la condena incorporan un tipo de sentido que en otras circunstancias no tendrían. Segundo, ahora estamos en condiciones de explicar las relaciones peculiares que los animales racionales tienen con el tiempo. Si puedo organizar el tiempo y planificar para el día de mañana, es porque el mismo yo que hace los planes existirá en el futuro para llevarlos a la práctica.

VI. Conclusión

En este capítulo me he ocupado sobre todo de dos problemas, en primer lugar los criterios de la identidad personal o, en otras palabras, qué hay en una persona que hace de ella la misma a través del tiempo y los cambios. En segundo lugar, traté de exponer un argumento para demostrar que, si bien Hume tenía razón al sostener que no hay yo alguno como objeto de nuestras

experiencias, existe, no obstante, una exigencia formal o lógica de postular un yo como algo sumado a ellas con el objeto de entender su carácter. En lo concerniente al planteamiento del argumento, no estoy insatisfecho. Pero sí lo estoy, y mucho, por el hecho de que a mi juicio no va lo bastante lejos y realmente no sé cómo completarlo. Tengo dos preocupaciones conexas. Primero, la dificultad subyacente con respecto a Hume era su concepción atomista de la experiencia. El filósofo creía que las experiencias siempre llegaban a nosotros en unidades discretas que llamaba "impresiones" e "ideas". Sabemos, sin embargo, que no es así. Como he tratado de poner de relieve, sabemos que tenemos un campo consciente total y unificado y que en él nuestras experiencias se organizan tanto en cualquier momento dado como a lo largo del tiempo en estructuras muy ordenadas y complejas. Los psicólogos *gestálticos* nos proporcionaron una multitud de pruebas del carácter holista y no atomista de nuestras experiencias perceptivas. La segunda preocupación es que no sé cómo explicar el hecho de que un importante rasgo de nuestras experiencias sea lo que podríamos denominar "sentido del yo". Una manera de expresarlo es decir que existe decididamente algo consistente en sentir que uno es uno mismo. Y un modo de verlo es tratar de imaginar cómo será ser alguien totalmente diferente. Imagine el lector que es Adolf Hitler, Napoleón o George Washington. Y al realizar este ejercicio imaginativo es importante no hacer trampas y no imaginarse en la situación de Adolf Hitler, etc.; no hay que pensarse en el papel de Hitler, sino tratar de imaginar cómo es ser Hitler. Si el lector lo hace, creo que advertirá que imagina una experiencia muy diferente de la experiencia habitual en la que tiene una idea de su yo como este yo y no otro. Pero

la existencia del sentido del yo no resuelve, desde luego, el problema de la identidad personal. Admitiendo la existencia de algo consistente en sentirse uno mismo, eso no basta para garantizar que quienquiera que tenga la experiencia debe ser idéntico a mí, porque es muy posible que muchas otras personas vivan esta misma experiencia de idéntico tipo que yo llamo "sentido de lo que es ser yo". Mi sentido del yo existe sin lugar a dudas, pero no resuelve el problema de la identidad personal y hasta ahora tampoco da carnadura al requisito puramente formal que juzgué necesario para complementar la descripción de Hume a fin de explicar la posibilidad de la acción libre racional. Por lo tanto, aunque este capítulo es un comienzo de la discusión del yo, no es más que eso: un comienzo.

EPÍLOGO

LA FILOSOFÍA Y LA COSMOVISIÓN CIENTÍFICA

He completado la tarea que me propuse en el primer capítulo. Intenté presentar una descripción de la mente que situara los fenómenos mentales como parte del mundo natural. Nuestra presentación de la mente en todos sus aspectos –conciencia, intencionalidad, libre albedrío, causalidad mental, percepción, acción intencional, etc.– es naturalista en este sentido: en primer lugar, trata los fenómenos mentales como parte de la naturaleza. Debemos concebir la conciencia y la intencionalidad en cuanto partes tan legítimas del mundo natural como la fotosíntesis o la digestión. Segundo, el aparato explicativo que usamos para proponer una descripción causal de los fenómenos mentales es un aparato que necesitamos para explicar la naturaleza en general. Intentamos explicar los fenómenos mentales en un nivel biológico y no, digamos, en el nivel de la física subatómica. La razón es que la conciencia y otros fenómenos mentales son fenómenos biológicos; son el producto de procesos biológicos y específicos de ciertos organismos biológicos. Esto no significa, desde luego, negar que nuestra mente es modelada por nuestra cultura. Pero la cultura no se opone a la biología; antes bien, es la forma adoptada por esta en diferentes comunidades. Una cultura puede diferir de otra, pero las diferencias tienen sus límites. Cada una debe ser una expresión de la comunidad biológica subyacente de la especie humana. No podría haber un conflicto a largo plazo entre la naturaleza y la cultura, porque si lo hu-

biera, la primera siempre ganaría y la segunda siempre perdería.

La gente habla a veces de la "cosmovisión científica" como si fuera una visión entre otras de cómo son las cosas y pudiera haber toda clase de cosmovisiones; la "ciencia", entonces, nos propondría una más de ellas. En algún aspecto es así; pero en otros la idea es engañosa y sugiere, en rigor, algo falso. Es posible observar la misma realidad con diferentes intereses en mente. Hay un punto de vista económico, un punto de vista estético, un punto de vista político, etc., y el punto de vista de la investigación científica es, en este sentido, uno más entre otros. Sin embargo, una manera de interpretar esta concepción sugiere que la ciencia designa un tipo específico de ontología, como si hubiera una realidad científica diferente, por ejemplo, de la realidad del sentido común. Creo que esto es un profundo error. La idea implícita en este libro, que ahora quiero explicitar, es que la ciencia no designa un dominio ontológico, sino un conjunto de métodos para indagar en todos los terrenos que admiten una investigación sistemática. La presencia de un electrón en los átomos de hidrógeno, por ejemplo, fue descubierta por medio de algo denominado "método científico", pero una vez descubierto, el hecho no es propiedad de la ciencia; es íntegramente de propiedad pública. Es un hecho como cualquier otro. Así, si nos interesan la realidad y la verdad, no hay en rigor nada que pueda llamarse "realidad científica" o "verdad científica". Sólo existen los hechos que conocemos. El desconocimiento de estos factores generó una indecible confusión en la filosofía. Por eso suele haber debates, por ejemplo, sobre la realidad de las entidades postuladas por la ciencia. Pero esas entidades o bien existen o bien no existen. La concepción que

tengo del tema es la siguiente: la existencia de un solo electrón en los átomos de hidrógeno es un dato similar al hecho de que yo tenga una sola nariz. La única diferencia es que, por razones evolutivas bastante accidentales, no necesito ayuda profesional para descubrir que tengo una sola nariz, mientras que, dadas nuestra estructura y la estructura de los átomos de hidrógeno, hace falta mucha pericia profesional para descubrir cuántos electrones hay en un átomo de ese elemento.

El mundo científico no existe. Lo que existe es simplemente el mundo, y el objetivo de nuestro afán es describir su funcionamiento y nuestra situación en él. Por lo que sabemos, sus principios más fundamentales son expuestos por la física atómica y, en cuanto al pequeño fragmento de ese mundo que más nos concierne, por la biología evolutiva. Los dos principios básicos de los cuales depende cualquier investigación como la que yo he emprendido son, primero, la noción de que las entidades más fundamentales de la realidad son las descritas por la física atómica, y segundo, que nosotros, como bestias biológicas, somos el producto de largos períodos de evolución, extendidos, quizá, durante cinco mil millones de años. Ahora bien, una vez que aceptamos estos puntos, que no sólo se refieren a la ciencia sino al funcionamiento del mundo, algunos de los interrogantes sobre la mente humana admiten respuestas filosóficas bastante simples, aunque esto no implica que las respuestas neurobiológicas lo sean igualmente.

No vivimos en varios y ni siquiera en dos mundos diferentes, uno físico y otro mental, un mundo científico y un mundo del sentido común. Por el contrario, hay un solo mundo, el mundo donde todos vivimos, y es preciso explicar nuestra existencia como parte de él.

SUGERENCIAS PARA MÁS LECTURAS

1. Una docena de problemas de filosofía de la mente

Descartes, R., *The Philosophical Writings of Descartes*, traducción de J. Collingham, R. Stoothoff y D. Murdoch, dos volúmenes, Cambridge, Cambridge University Press, 1985, vol. 2, en especial *Meditations on First Philosophy*, "Second meditation", pp. 16-23 y "Sixth meditation", pp. 50-62, y *Objections and Replies*, en especial "Author's replies to the fourth set of objections", pp. 154-162 [traducción española: *Meditaciones metafísicas con objeciones y respuestas*, Madrid, Alfaguara, 1977].

Hay una serie de introducciones generales a la filosofía de la mente, entre las que cabe mencionar las siguientes:

Armstrong, D. M., *The Mind-Body Problem, An Opinionated Introduction*, Boulder (Colo.), Westview Press, 1999.

Churchland, P. M., *Matter and Consciousness*, Cambridge (Mass.), MIT Press, 1988 [traducción española: *Materia y conciencia: introducción contemporánea a la filosofía de la mente*, Barcelona, Gedisa, 1999].

Heil, J., *Philosophy of Mind*, Londres y Nueva York, Routledge, 1998.

Jacquette, D., *Philosophy of Mind*, Englewood Cliffs, Nueva Jersey, Prentice Hall, 1994.

Kim, J., *The Philosophy of Mind*, Boulder (Colo.), Westview Press, 1998.

Lyons, W., *Matters of the Mind*, Nueva York, Routledge, 2001.

También hay varias compilaciones de artículos sobre la filosofía de la mente, entre ellas:

Block, N. (comp.), *Readings in Philosophy of Psychology*, vol. 1, Cambridge (Mass.), Harvard University Press, 1980.

Chalmers, D. (comp.), *Philosophy of Mind: Classical and Contemporary Readings*, Nueva York, Oxford University Press, 2002.

Heil, J. (comp.), *Philosophy of Mind: A Guide and Anthology*, Oxford, Oxford University Press, 2004.

Lycan, W. (comp.), *Mind and Cognition: A Reader*, Cambridge (Mass.), Blackwell, 1990.

O'Connor, T. y D. Robb (comps.), *Philosophy of Mind: Contemporary Readings*, Londres y Nueva York, Routledge, 2003.

Rosenthal, D. M. (comp.), *The Nature of Mind*, Nueva York, Oxford University Press, 1991.

2. El giro hacia el materialismo

Las siguientes selecciones presentan la mayor parte de los argumentos básicos examinados en este capítulo:

Armstrong, D. M., *A Materialist Theory of the Mind*, Londres, Routledge, 1993.

Block, N., "Troubles with Functionalism", en D. Wade Savage (comp.), *Perception and Cognition: Issues in the Foundations of Psychology*, Minneápolis, University of Minnesota Press, 1978, col. "Minnesota Studies in the Philosophy of Science", vol. 9, pp. 261-325, reeditado en N. Block (comp.), *Readings in Philosophy of Psychology*, vol. 1, Cambridge (Mass.), Harvard University Press, 1980, pp. 268-305.

Borst, C. (comp.), *The Mind/Brain Identity Theory*, Nueva York, St. Martin's Press, 1970.

Churchland, P. M., "Eliminative Materialism and the Propositional Attitudes", en D. M. Rosenthal (comp.), *The Nature of Mind*, Nueva York, Oxford University Press, 1991, pp. 601-612 [traducción española: "El materialismo eliminativo y las actitudes proposicionales", en Eduardo Rabossi (comp.), *Filosofía de la mente y conciencia cognitiva*, Barcelona, Paidós, 1995, pp. 43-68].

Crane, T., *The Mechanical Mind*, segunda edición, Londres, Routledge, 2003.

Davidson, D., "Mental Events", en *Essays on Actions and Events*, Nueva York, Oxford University Press, 1980, pp. 207-227 [traducción española: "Sucesos mentales", en *Ensayos sobre acciones y sucesos*, Barcelona y México, Crítica/Instituto de Investigaciones Filosóficas de la UNAM, 1995].

Feigl, H., "The 'Mental' and the 'Physical'", en H. Feigl, M. Scriven y G. Maxwell (comps.), *Concepts, Theories and the Mind-Body Problem*, Minneápolis, University of Minnesota Press, 1958, col. "Minnesota Studies in the Philosophy of Science", vol. 2.

Haugeland, J. (comp.), *Mind Design: Philosophy, Psychology, Artificial Intelligence*, Cambridge (Mass.), MIT Press, 1982, col. "A Bradford Book".

Hempel, C., "The Logical Analysis of Psychology", en N. Block (comp.), *Readings in Philosophy of Psychology*, vol. 1, Cambridge (Mass.), Harvard University Press, 1980.

Lewis, D., "Psychophysical and Theoretical Identifications" y "Mad Pain and Martian Pain", en N. Block (comp.), *Readings in Philosophy of Psychology*, vol. 1, Cambridge (Mass.), Harvard University Press, 1980, pp. 207-215 y 216-222, respectivamente.

McDermott, D. V., *Mind and Mechanism*, Cambridge (Mass.), MIT Press, 2001.

Nagel, T., "Armstrong on the Mind", en N. Block (comp.), *Readings in Philosophy of Psychology*, vol. 1, Cambridge (Mass.), Harvard University Press, 1980.

Place, U. T., "Is Consciousness a Brain Process?", *British Journal of Psychology*, 47, primera parte, 1956, pp. 44-50.

Putnam, H., "The Nature of Mental States", en N. Block (comp.), *Readings in Philosophy of Psychology*, vol. 1, Cambridge (Mass.), Harvard University Press, 1980 [traducción española: *La naturaleza de los estados mentales*, México, Instituto de Investigaciones Filosóficas de la UNAM, 1981].

Ryle, G., *The Concept of Mind*, Londres, Hutchinson, 1949 [traducción española: *El concepto de lo mental*, Buenos Aires, Paidós, 1967].

Searle, J. R., *The Rediscovery of the Mind*, Cambridge (Mass.), MIT Press, 1992 [traducción española: *El redescubrimiento de la mente*, Barcelona, Crítica, 1996].

Smart, J. J. C., "Sensations and Brain Processes", en D. M. Rosenthal (comp.), *The Nature of Mind*, Nueva York, Oxford University Press, 1991, pp. 169-176.

Turing, A., "Computing Machinery and Intelligence", *Mind*, 59, 1950, pp. 433-460.

3. Argumentos contra el materialismo

Block, N., "Troubles with Functionalism", en D. Wade Savage (comp.), *Perception and Cognition: Issues in the Foundations of Psychology*, Minneápolis, University of Minnesota Press, 1978, col. "Minnesota Studies in the Philosophy of Science", vol. 9, pp. 261-325, reeditado en N. Block (comp.), *Readings in Philosophy of Psychology*, vol. 1, Cambridge (Mass.), Harvard University Press, 1980, pp. 268-305.

Jackson, F., "What Mary Didn't Know", *Journal of Philosophy*, 83, 1986, pp. 291-295 [traducción española: "Lo que María no sabía", en Obeth Hansberg y Maite Ezcurdia (comps.), *La naturaleza de la experiencia*, 1, *Sensaciones*, México, Instituto de Investigaciones Filosóficas de la UNAM, 2003]; véase también "Epiphenomenal Qualia", *Philosophical Quarterly*, 32, 1986, pp. 127-136 [traducción española: "Qualia epifenoménicos", en *ibid.*]

Kripke, S. A., *Naming and Necessity*, Cambridge (Mass.), Harvard University Press, 1980 [traducción española: *El nombrar y la necesidad*, México, UNAM, 1996]; extractos en D. Chalmers (comp.), *Philosophy of Mind: Classical and Contemporary Readings*, Nueva York, Oxford University Press, 2002, pp. 329-332.

McGinn, C., "Anomalous Monism and Kripke's Cartesian Intuitions", en N. Block (comp.), *Readings in Philosophy of Psychology*, vol. 1, Cambridge (Mass.), Harvard University Press, 1980, pp. 156-158.

Nagel, T., "Armstrong on the Mind", en N. Block (comp.), *Readings in Philosophy of Psychology*, vol.

1, Cambridge (Mass.), Harvard University Press, 1980, pp. 200-206.

Nagel, T., *The View from Nowhere*, Nueva York, Oxford University Press, 1986 [traducción española: *Una visión de ningún lugar*, Madrid, Fondo de Cultura Económica, 1996].

Nagel, T., "What Is It Like to Be a Bat?", *Philosophical Review*, 83, 1974, pp. 435-450, reeditado en D. Chalmers (comp.), *Philosophy of Mind: Classical and Contemporary Readings*, Nueva York, Oxford University Press, 2002 [traducción española: "¿Cómo es ser un murciélago?", en Obeth Hansberg y Maite Ezcurdia (comps.), *La naturaleza de la experiencia*, 1, *Sensaciones*, México, Instituto de Investigaciones Filosóficas de la UNAM, 2003].

Searle, J. R., "Minds, Brains and Programs", *Behavioral and Brain Sciences*, 3, 1980, pp. 417-424, reeditado en T. O'Connor y D. Robb (comps.), *Philosophy of Mind: Contemporary Readings*, Londres y Nueva York, Routledge, 2003, pp. 332-352 [traducción española: "Mentes, cerebros y programas", en Margaret A. Boden (comp.), *Filosofía de la inteligencia artificial*, México, Fondo de Cultura Económica, 1994].

Searle, J. R., *The Rediscovery of the Mind*, Cambridge (Mass.), MIT Press, 1992 [traducción española: *El redescubrimiento de la mente*, Barcelona, Crítica, 1996].

4. La conciencia, primera parte

Hay una multitud de trabajos recientes sobre la conciencia, incluidos algunos de este autor. A continuación, una muestra representativa.

Chalmers, D. *The Conscious Mind: In Search of a Fundamental Theory*, Oxford, Oxford University Press, 1996 [traducción española: *La mente consciente: en busca de una teoría fundamental*, Barcelona, Gedisa, 1999].

Dennett, D., *Consciousness Explained*, Boston, Little Brown, 1991 [traducción española: *La conciencia explicada: una teoría interdisciplinar*, Barcelona, Paidós, 1995].

McGinn, C., *The Problem of Consciousness: Essays toward a Resolution*, Cambridge (Mass.), Basil Blackwell, 1991.

Nagel, T., *The View from Nowhere*, Nueva York, Oxford University Press, 1986 [traducción española: *Una visión de ningún lugar*, Madrid, Fondo de Cultura Económica, 1996].

O'Shaughnessy, B., *Consciousness and the World*, Oxford, Oxford University Press, 2000.

Searle, J. R., *The Mystery of Consciousness*, Nueva York, New York Review of Books, 1997 [traducción española: *El misterio de la conciencia*, Barcelona, Paidós, 2000].

Searle, J. R., *The Rediscovery of the Mind*, Cambridge (Mass.), MIT Press, 1992 [traducción española: *El redescubrimiento de la mente*, Barcelona, Crítica, 1996].

Siewert, C., *The Significance of Consciousness*, Princeton, Princeton University Press, 1998.

Tye, M., *Ten Problems of Consciousness*, Cambridge (Mass.), MIT Press, 1995.

También hay una enorme antología (más de ochocientas páginas) de artículos sobre la conciencia:

Block, N., O. Flanagan y G. Guzeldere (comps.), *The Nature of Consciousness: Philosophical Debates*, Cambridge (Mass.), MIT Press, 1997.

Al final de "La conciencia, segunda parte" se mencionarán lecturas más orientadas hacia la neurobiología.

5. La conciencia, segunda parte

Los enfoques neurobiológicos de la conciencia son variados. Entre ellos:

Crick, F., *The Astonishing Hypothesis*, Nueva York, Scribner's, 1994 [traducción española: *La búsqueda científica del alma: una revolucionaria hipótesis para el siglo XXI*, Madrid, Debate, 1995].

Damasio, A. R., *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, Nueva York, Harcourt Brace & Co., 1999 [traducción española: *La sensación de lo que ocurre: cuerpo y emoción en la construcción de la conciencia*, Madrid, Debate, 2001].

Edelman, G., *The Remembered Present*, Nueva York, Basic Books, 1989.

Koch, C., *The Quest for Consciousness: A Neurobiological Approach*, Englewood (Colo.), Roberts and Co., 2004.

Llinás, R., *I of the Vortex: From Neurons to Self*, Cambridge (Mass.), MIT Press, 2001 [traducción española: *El cerebro y el mito del yo*, Bogotá, Norma, 2003].

Searle, J. R., "Consciousness", *Annual Review of Neuroscience*, 23, 2000, reeditado en J. R. Searle, *Consciousness and Language*, Cambridge, Cambridge University Press, 2002. (Este artículo contiene una extensa bibliografía de las investigaciones neurobiológicas actuales sobre la conciencia.)

6. La intencionalidad

Burge, T., "Individualism and the Mental", en P. A. French, T. E. Uehling, Jr. y H. K. Wettstein, *Studies in Metaphysics*, Minneápolis, University of Minnesota Press, 1979, col. "Midwest Studies in Philosophy", vol. 4.

Fodor, J., "Meaning and the World Order", en *Psychosemantics*, Cambridge (Mass.), MIT Press, 1988, capítulo 4, reeditado en T. O'Connor y D. Robb (comps.), *Philosophy of Mind: Contemporary Readings*, Londres y Nueva York, Routledge, 2003 [traducción española: *Psicosemántica: el problema del significado en la filosofía de la mente*, Madrid, Tecnos, 1994].

Putnam, H., "The Meaning of 'Meaning'", en K. Gunderson (comp.), *Language, Mind, and Knowledge*, Minneápolis, University of Minnesota Press, 1975, col. "Minnesota Studies in the Philosophy of Science", vol. 7, pp. 131-193 [traducción española: "El significado de 'significado'", en Luis Valdés Villanueva (comp.), *La búsqueda del significado: lecturas de filosofía del lenguaje*, Madrid, Tecnos, 1995].

Searle, J. R., *Intentionality: An Essay in the Philosophy of Mind*, Cambridge, Cambridge University Press, 1983 [traducción española: *Intencionalidad: un ensayo en filosofía de la mente*, Madrid, Tecnos, 1992].

7. La causación mental

Davidson, D., "Actions, Reasons and Causes", en *Essays on Actions and Events*, Nueva York, Oxford University Press, 1980 [traducción española: "Acciones, razones y causas", en *Ensayos sobre acciones y sucesos*, Barcelona y México, Crítica/Instituto de Investigaciones Filosóficas de la UNAM, 1995].

Heil, J. y A. Mele (comps.), *Mental Causation*, Oxford, Clarendon Press, 1993.

Kim, J., *Mind in a Physical World: An Essay on the Mind-Body Problem and Causation*, Cambridge (Mass.), MIT Press, 1998.

Searle, J. R., *Intentionality: An Essay in the Philosophy of Mind*, Cambridge, Cambridge University Press, 1983 [traducción española: *Intencionalidad: un ensayo en filosofía de la mente*, Madrid, Tecnos, 1992].

8. El libre albedrío

Hay una antología de artículos sobre el libre albedrío en:

Watson, G. (comp.), *Free Will*, segunda edición, Oxford, Oxford University Press, 2003.

A continuación, algunos libros recientes:

Kane, R., *The Significance of Free Will*, Oxford, Oxford University Press, 1996.

Searle, J. R., *Rationality in Action*, Cambridge (Mass.), MIT Press, 2001.

Smilansky, S., *Free Will and Illusion*, Oxford, Oxford University Press, 2002.

Wegner, D. N., *The Illusion of Conscious Will*, Cambridge (Mass.), MIT Press, 2003.

Wolf, S., *Freedom with Reason*, Oxford, Oxford University Press, 1994.

9. El inconsciente y la explicación del comportamiento

Freud, S., "A Note on the Unconscious in Psychoanalysis" (1912), en *Collected Papers*, traducción de J. Riviere, vol. 4, Nueva York, Basic Books, 1959, pp. 22-29 [traducción española: "Nota sobre el concepto de inconsciente en psicoanálisis", en *Obras completas*, vol. 12, Buenos Aires, Amorrortu editores, 1980].

Freud, S., "The Unconscious" (1915), en *Collected Papers*, traducción de J. Riviere, vol. 4, Nueva York, Basic Books, 1959, pp. 98-136 [traducción española: "Lo inconsciente", en *Obras completas*, vol. 14, Buenos Aires, Amorrortu editores, 1979].

Searle, J. R., *The Rediscovery of the Mind*, Cambridge (Mass.), MIT Press, 1992, capítulo 7 [traducción española: *El redescubrimiento de la mente*, Barcelona, Crítica, 1996].

Searle, J. R., *Rationality in Action*, Cambridge (Mass.), MIT Press, 2001.

10. La percepción

El ataque clásico contra las teorías realistas de la percepción se encontrará en:

Berkeley, G., *Principles of Human Knowledge*, edición establecida por J. Dancy, Oxford, Oxford University Press, 1998 [traducción española: *Tratado sobre los principios del conocimiento humano*, Madrid, Alianza, 1984]. Véase también Berkeley, G., *Three Dialogues between Hylas and Philonous*, edición establecida por C. Turbayne, Indianápolis, Bobbs-Merrill Educational Publishing, 1985 [traducción española: *Tres diálogos entre Hilas y Filonús*, Madrid, Espasa-Calpe, 1996].

Se encontrará una formulación moderna de las teorías de los datos de los sentidos en:

Ayer, A. J., *The Foundations of Empirical Knowledge*, Londres, Macmillan, 1953.

Para una crítica de esa misma teoría, véase:

Austin, J. L., *Sense and Sensibilia*, edición establecida por G. J. Warnock, Oxford, Clarendon Press, 1962 [traducción española: *Sentido y percepción*, Madrid, Tecnos, 1981].

Se hallará una descripción de la intencionalidad de la percepción en:

Searle, J. R., *Intentionality: An Essay in the Philosophy of Mind*, Cambridge, Cambridge University Press, 1983, capítulo 2 [traducción española: *Intencionalidad: un ensayo en filosofía de la mente*, Madrid, Tecnos, 1992].

11. El yo

La formulación clásica del escepticismo con respecto al yo está en:

Hume, D., *A Treatise of Human Nature*, edición establecida por L. A. Selby-Bigge, Oxford, Clarendon Press, 1951, libro I, cuarta parte, sección VI, sobre la identidad personal, pp. 251-263, así como en el apéndice, pp. 623-939 [traducción española: *Tratado de la naturaleza humana*, Barcelona, Orbis, 1981].

La concepción de Locke se encontrará en:

Locke, J., *Essay Concerning Human Understanding*, Londres, Routledge, 1894, en especial el capítulo 27, "Of Identity and Diversity" [traducción española: *Ensayo sobre el entendimiento humano*, México, Fondo de Cultura Económica, 1992].

Otras obras sobre problemas planteados en este capítulo:

Parfit, D., *Reasons and Persons*, Oxford, Oxford University Press, 1986.

Searle, J. R., *Rationality in Action*, Cambridge (Mass.), MIT Press, 2001, sobre todo el capítulo 3.

El siguiente libro es una colección de ensayos:

Perry, J. (comp.), *Personal Identity*, Berkeley y Los Angeles, University of California Press, 1975.