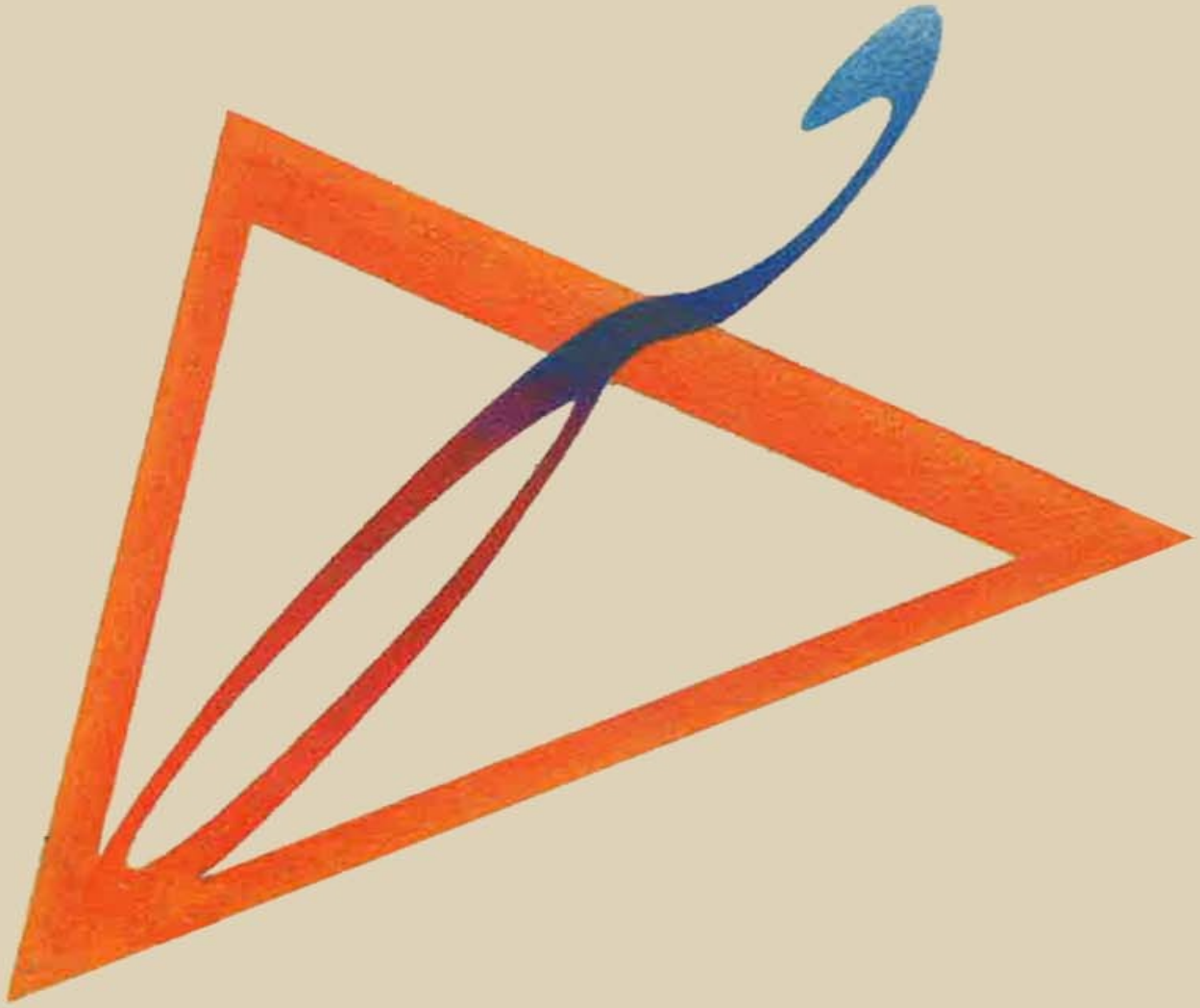
Filosofía de la mente

Una panorámica para la ciencia cognitiva

William Bechtel



El estudio de la mente ha sido, y continúa siendo, una de las preocupaciones centrales de los filósofos que, a lo largo de la historia, han enunciado toda una hueste de tesis sustantivas sobre su naturaleza y la de la actividad mental. El surgimiento en nuestra época de lo que ha dado en llamarse "ciencia cognitiva", lejos de constituir la piedra de toque para decidir entre las distintas posiciones en competición, obliga a tomar partido explícita o implícitamente respecto a ellas.

Este libro presenta, de una manera clara y comprensible para los no especialistas, las perspectivas filosóficas sobre la mente. Después de un recorrido histórico y de una consideración de las importantes relaciones entre la filosofía del lenguaje y la filosofía de la mente, se examina el problema de la intencionalidad y las distintas estrategias filosóficas para explicarla. A continuación se presenta el problema mente/cuerpo: el dualismo y las distintas versiones del conductismo y del materialismo. El último capítulo está dedicado a la exposición y crítica del análisis dominante de los eventos mentales en la filosofía de la mente actual: el funcionalismo.

eBooks con estilo

William Bechtel

Filosofía de la mente

Una panorámica para la ciencia cognitiva

ePUB v1.0
botasdesieteleguas 13.09.2011

Título original: *Philosophy of Mind. An Overview for Cognitive Science*.
Lugar y año de la edición original: New Jersey, USA, 1988
Versión castellana: Luis Ml. Valdés Villanueva
Editiorial: EDITORIAL TECNOS, S.A.
Lugar y año de edición: Madrid, 1991
ISBN: 84-309-2038-2

A mi madre

PRÓLOGO

Como una de las diversas disciplinas que contribuyen a la ciencia cognitiva, la filosofía ofrece dos tipos de contribuciones. Por una parte, la *filosofía de la ciencia* proporciona una perspectiva metateórica de los propósitos de cualquier empresa científica, analizando cosas tales como las metas de la investigación científica y las estrategias empleadas para alcanzar esas metas. Por otra, la *filosofía de la mente* ofrece tesis sustantivas sobre la naturaleza de la mente y de la actividad mental. Aunque esas tesis no son típicamente resultado de la investigación empírica, han figurado a menudo subsiguientemente en las investigaciones empíricas efectivas de la ciencia cognitiva o de sus predecesoras. Puesto que los dos papeles que desempeña la filosofía de la ciencia cognitiva son completamente distintos, se introducen en dos volúmenes separados. Éste se centra en la filosofía de la mente, mientras que los problemas de filosofía de la ciencia se exploran en *Filosofía de la ciencia: Una panorámica para la ciencia cognitiva*.

La meta de este libro es proporcionar una amplia visión general de los problemas centrales de la filosofía de la mente y una introducción a la literatura profesional. Los filósofos han adoptado una gran variedad de posiciones respecto de los problemas que discuto y he intentado describir de la manera más simple posible las posiciones más prominentes. Me he propuesto también citar un amplio rango de artículos y libros filosóficos cuya consulta recomiendo al lector para desarrollar una comprensión más cabal de las distintas posiciones que han tomado los filósofos.

Comienzo con un capítulo que discute la metodología de la investigación filosófica a la vez que ofrece una visión general de las figuras más importantes de la historia de la filosofía cuyas ideas han sido influyentes en filosofía de la mente y en ciencia cognitiva de un modo general. A continuación, en el capítulo 2, discuto varias explicaciones del lenguaje que han sido desarrolladas por filósofos analíticos durante el siglo XX. Mente y lenguaje son obviamente fenómenos estrechamente relacionados, y las perspectivas desarrolladas en los análisis del lenguaje han influido en las explicaciones filosóficas de la mente. Hago, por tanto, repetidas referencias a este material en los capítulos siguientes. Los análisis filosóficos del lenguaje han tenido también una influencia considerable sobre el trabajo de otras disciplinas [11-12] de la ciencia cognitiva, incluyendo la lingüística y la psicología cognitiva. Muchos filósofos han considerado la intencionalidad como el rasgo distintivo de los fenómenos mentales. Los capítulos 3 y 4 están dedicados a exponer diferentes explicaciones que los filósofos han ofrecido de lo que es la intencionalidad y de cómo ha de ser considerada para distinguir la mente de otros fenómenos de la naturaleza. Algunos filósofos han considerado la intencionalidad como algo que diferencia tanto las mentes de las demás cosas de la naturaleza, que hace imposible el desarrollar una ciencia de la mente. Las afirmaciones de tales filósofos se discuten en el capítulo 3. El capítulo 4 está dedicado a examinar cierto número de intentos por parte de otros filósofos de mostrar cómo la intencionalidad puede surgir en el mundo natural y cómo la intencionalidad de los eventos mentales podría explicarse científicamente. Varios de esos intentos han sido directamente motivados por el trabajo reciente en ciencia cognitiva, y las respuestas apuntan a diferentes tipos de propósitos de investigación que la ciencia cognitiva ha de perseguir.

Quizás el problema más ampliamente discutido en filosofía de la mente durante los últimos tres siglos ha sido el *problema mente-cuerpo*. Este problema es un legado de Descartes y se han propuesto muchas

respuestas para él. En los capítulos 5 y 6 examino cierto número de esas respuestas y sus implicaciones para la ciencia cognitiva. El capítulo 5 comienza con un examen de diferentes formas de dualismo, dedicando atención primaria al dualismo de substancias. Esta posición considera las mentes como géneros de cosas totalmente diferentes de los cuerpos y, por consiguiente, parece rechazar para siempre la posibilidad de desarrollar explicaciones de la actividad mental usando estrategias de la ciencia natural. En ese capítulo discuto también el conductismo filosófico, uno de los primeros intentos sistemáticos de rechazar el dualismo. Aunque el conductismo filosófico y el conductismo en psicología tienen diferentes aspiraciones, ambos se oponen a usar modelos de procesamiento interno para explicar la conducta y son, por tanto, antitéticos respecto de los propósitos de la ciencia cognitiva.

El capítulo 6 examina cierto número de variedades de materialismo, que mantiene que los estados mentales son estados del cerebro. La Teoría de la Identidad como Tipo se desarrolló como respuesta al trabajo en las neurociencias que sugerían una correlación entre géneros de estados mentales y tipos de estados neurales. Se proponía que tener un cierto género de estado mental era justamente estar en un estado neural particular. La Teoría de la Identidad como Tipo es entonces completamente compatible con modelos de procesamiento interno [12-13] de cognición, pero liga esos modelos estrechamente a los de la neurociencia. Niega, por tanto, cualquier autonomía a las investigaciones de la ciencia cognitiva. El materialismo eliminativo simpatiza menos aún con la idea de una ciencia cognitiva autónoma, manteniendo que las teorías mentalistas deberían reemplazarse por teorías desarrolladas a partir de la neurociencia. Una tercera forma de materialismo, la Teoría de la Identidad como Instancia, es la solución que se ha intentado al problema mente-cuerpo que resulta más afín a la ciencia cognitiva. Mantiene que cada estado mental individual es también un estado del cerebro, pero niega que la taxonomía de los estados mentales corresponda a la taxonomía de los estados neurales. Así pues, permite que las explicaciones cognitivas de la conducta sean completamente independientes de las explicaciones neurales.

El cognitivismo ha planteado un problema especial a le ha sido el foco de gran parte de la obra reciente en filosofía de la mente. Al desarrollar los modelos de procesamiento interno, los cognitivistas intentan caracterizar los eventos mentales en términos de su eficacia causal. Una teoría filosófica llamada Funcionalismo intenta caracterizar este modo de identificar y clasificar eventos mentales. Esta teoría es el centro de atención del último capítulo. Introduzco varias versiones de Funcionalismo que se han desarrollado en filosofía de la mente y discuto también cierto número de objeciones que se han planteado contra el Funcionalismo. Concluyo describiendo una forma alternativa de Funcionalismo desarrollada en filosofía de la biología y muestro cómo proporciona un modo potencial más fructífero de clasificar los eventos mentales.

Para aquellos que no estén familiarizados previamente con la filosofía son convenientes algunos comentarios acerca de cómo enfocar el material filosófico. Aunque suele proclamarse ampliamente que las afirmaciones filosóficas no exigen evidencia empírica, este punto de vista es cada vez menos aceptado en la actualidad. Cierto número de tesis discutidas en filosofía de la mente se desarrollaron como análisis del trabajo empírico realizado en psicología empírica y otras ciencias cognitivas. Sin embargo, continúa siendo verdad que las afirmaciones filosóficas tienden a estar bastante apartadas de la evidencia empírica. Por tanto, en este tipo de afirmaciones tiende a haber mucho más lugar para discutir sobre sus correspondientes virtudes que en el caso de disciplinas donde la evidencia empírica está

fácilmente a la mano.

Al considerar los puntos de vista discutidos en este libro, el lector debe recordar el carácter controvertido y argumentativo de la investigación filosófica. Más bien que aceptar o rechazar simplemente un punto de vista, el lector debe considerar los géneros posibles de argumento [13-14] que la mente puede ofrecer a favor o en contra de ellos. El lector entra, por tanto, dentro del argumento mismo, y no permanece como un observador pasivo. Aunque los esfuerzos acumulados de los filósofos para hacer frente a esos problemas proporcionan un recurso para cualquiera que desee abordarlos, los problemas no son la prerrogativa exclusiva de los filósofos y debe animarse a los científicos para que se ocupen de discutir los problemas mismos y alcanzar sus propias conclusiones.

AGRADECIMIENTOS

Mientras escribía este texto he recibido ayuda y apoyo de gran número de personas e instituciones. En primer lugar, gracias a Larry Erlbaum por invitarme a escribir este texto. Aunque no fue un proyecto tan fácil como parecía cuando me invitó a hacerlo, he aprendido mucho de él. Debo también un agradecimiento especial a Andrew Ortony por su valioso consejo editorial y por sus comentarios. Jim Frame fue mi ayudante de investigación durante la mayor parte del tiempo en el que estuve escribiendo este texto y me proporcionó una ayuda inestimable, particularmente organizando y coordinando materiales bibliográficos. Adele Abrahamsen, Robert McCauley, Donald Norman. Richard Robinson y Douglas Winblad leyeron varias versiones de este texto y me ofrecieron comentarios substanciales por los que estoy muy agradecido. He usado versiones preliminares de este texto en mi curso de Filosofía de la Psicología en la Georgia State University en el semestre de otoño de 1985, y estoy muy agradecido a los estudiantes de ese curso por haberme retroalimentado de modo tan útil. Finalmente, una beca de investigación de la Georgia State University proporcionó un apoyo esencial para desarrollar el texto, y así lo reconozco con mi agradecimiento. [14-15]

1. ALGUNAS PERSPECTIVAS SOBRE LA FILOSOFÍA DE LA MENTE

1.1 INTRODUCCIÓN: ¿QUÉ ES LA FILOSOFÍA DE LA MENTE?

Este libro está dedicado a introducir los problemas básicos de la filosofía de la mente para aquellos que practican otras disciplinas de la ciencia cognitiva: psicología cognitiva, inteligencia artificial, neurociencia cognitiva, lingüística teórica y antropología cognitiva. Los filósofos se interesaron por el carácter de la mente mucho antes de que surgieran esas disciplinas empíricas. Se planteaban cuestiones como las siguientes: ¿Cuáles son los rasgos distintivos de las mentes? ¿Cómo se deberían caracterizar los estados mentales? ¿Cómo se relacionan las mentes con los cuerpos físicos? ¿Cómo son capaces las mentes de aprender cosas sobre el mundo físico? En los capítulos que siguen de este libro se examina un variado conjunto de respuestas que los filósofos han ofrecido a esas y otras preguntas. Antes de volver la vista hacia las concepciones particulares que los filósofos han avanzado, es útil, sin embargo, ofrecer una perspectiva de las investigaciones filosóficas de esos problemas.

Hay dos cuestiones que los científicos cognitivos no filosóficamente entrenados plantearán muy probablemente sobre la filosofía de la mente, a saber: *a*) ¿qué metodología emplean los filósofos para analizar los fenómenos mentales? y *b*) ¿cómo se relacionan los esfuerzos de los filósofos con las investigaciones llevadas a cabo en otras disciplinas de la ciencia cognitiva? Planteo estos dos problemas en la primera sección del capítulo y, a continuación, ofrezco una visión general de alguna de las principales tradiciones históricas en filosofía que proporcionan tanto los orígenes de muchas ideas que ahora son influyentes en ciencia cognitiva, como el trasfondo del pensamiento filosófico contemporáneo sobre estos asuntos.

Por lo que respecta a la metodología, la filosofía se distingue de otras disciplinas de la ciencia cognitiva en que no tiene su propia base empírica distintiva[1]. Los filósofos distinguen a menudo entre el conocimiento [15-16] *a priori*, que puede descubrirse sin investigación empírica, y el conocimiento *a posteriori*, que descansa sobre resultados empíricos. Muchos filósofos han pensado que pueden establecerse *a priori* verdades importantes sobre la mente. Mantienen que esas verdades pueden establecerse simplemente razonando sobre cómo ha de ser la mente o analizando la estructura de nuestro lenguaje mediante el cual hablamos sobre las mentes. Otros filósofos, aunque mantienen que sus afirmaciones son en última instancia *a posteriori*, han tratado de establecer verdades sobre la mente extrayendo algunas de las consecuencias lógicas de los resultados que los científicos han obtenido mediante investigación empírica.

Dentro de la filosofía, las discusiones sobre la naturaleza de la mente se producen generalmente en dos subapartados: la epistemología y la metafísica. La epistemología, que busca definir qué es el conocimiento y determinar cómo se obtiene, se interesa por aquellos procesos por medio de los que la mente es capaz de reunir conocimiento. La metafísica se ha caracterizado tradicionalmente como el estudio de los principios básicos del universo y de sus orígenes. La ontología, un subapartado de la

metafísica, se interesa por identificar y caracterizar los géneros de cosas que existen en el mundo[2]. Es precisamente en este apartado donde se estudia el carácter de la mente. Una porción del trabajo contemporáneo en ontología está estrechamente enlazada con los resultados de las investigaciones científicas y analiza qué géneros de objetos suponen esas ciencias que existen. Los filósofos se han interesado por asuntos tales como los criterios mediante los cuales determinamos si las entidades teóricas postuladas por las ciencias (tales como los quarks o los estados mentales) existen realmente o son simplemente ficciones útiles para hacer ciencia. Quine (1969a) avanzó la máxima (con la que no todos están de acuerdo) de que lo que consideramos que existe son las entidades postuladas en nuestras teorías científicas. El enfoque de Quine enlaza estrechamente la investigación de problemas metafísicos con los trabajos de la ciencia empírica, pero queda en pie la cuestión de cuándo deberíamos aceptar que una teoría científica proporciona un [16-17] enfoque adecuado de la naturaleza. Quine piensa que las teorías que pretenden hablar sobre estados mentales no son teorías científicas aceptables (véase el capítulo 3).

La mayor parte de los filósofos de hoy en día mantendrían que la ciencia empírica es relevante para las discusiones, tanto epistemológicas como ontológicas, sobre la mente pero con todo mantienen que los problemas filosóficos son distintos de los problemas empíricos que se plantean en otras disciplinas de la ciencia cognitiva. Generalmente, se piensa que la distinción es un resultado del hecho de que la filosofía se interesa por problemas conceptuales fundamentales. Tales problemas tienen que ver con la adecuación de una armazón teórica particular para acomodar rasgos de estados mentales como su intencionalidad (capítulos 3 y 4) o su carácter efectivo o cualitativo (capítulo 7). Éstos son problemas para los que simplemente no podemos diseñar experimentos empíricos. Por consiguiente, los intentos de responder a ellos involucran a menudo argumentos complejos que nos llevan muy lejos de los resultados empíricos.

El hecho de que las afirmaciones filosóficas estén tan separadas de la investigación empírica plantea un desafío a cualquiera que vuelva la vista hacia las investigaciones filosóficas desde su entrenamiento en la investigación experimental. Para evaluar una afirmación filosófica ha de seguirse a menudo la complicada cadena de razonamiento que se ofrece para apoyar la afirmación. Esto, no obstante, no pretende disuadir a los no involucrados de que entren en el ruedo filosófico. Es más: tal participación es muy de agradecer; uno de los beneficios que los filósofos pueden obtener de la participación en el racimo de investigación interdisciplinar de la ciencia cognitiva es el aprender nuevas perspectivas sobre la mente de otros científicos cognitivos.

Todo lo que se requiere para que el no filósofo logre involucrarse en la filosofía de la mente es que comience a hacer frente a los problemas. Esto significa convertirse en un participante activo en los debates ofreciendo argumentos a favor o en contra de las diferentes posiciones. No basta simplemente el volverse hacia los filósofos como autoridades y citar lo que un filósofo particular ha dicho como respuesta a una de esas cuestiones fundamentales. Dado que los puntos de vista filosóficos dependen de una larga cadena de argumentación frecuentemente son controvertidos. Diferentes filósofos mantienen una variedad de puntos de vista diferentes sobre esos problemas. Esto resultará evidente a medida que consideremos varios problemas en los capítulos siguientes. Más bien que aceptar simplemente una autoridad, es necesario explorar los problemas y evaluar los argumentos avanzados para las afirmaciones en competición. Sobre esta [17-18] base se puede esperar el tomar una decisión racional sobre qué posición aceptar[3].

Los no filósofos, una vez reconocida la naturaleza controvertida de las afirmaciones filosóficas, deciden algunas veces que tales cuestiones fundamentales no pueden ser resueltas. Se forman el punto de vista de que hay simplemente una variedad decisiones distintas y que no importa mucho cuál se acepte. El hecho de que los filósofos hayan estado planteando alguna de esas cuestiones desde hace 2.500 años y que aún no estén de acuerdo en cómo responderlas parecería proporcionar un buen apoyo para tal pretensión. Pero lo que esta afirmación no logra reconocer es que existe a menudo una estrecha interacción entre las afirmaciones filosóficas y los esfuerzos de la investigación empírica de tal manera que aquellos que toman parte en una investigación empírica suponen frecuentemente, de modo consciente o inconsciente, una postura filosófica particular. Históricamente, esas conexiones pueden mostrarse en la historia de la física o de la biología, pero basta considerar aquí algunas maneras en que los puntos de vista filosóficos han tenido o están teniendo un amplio impacto sobre la ciencia cognitiva.

El enfoque cognitivo de los fenómenos mentales, que unifica el trabajo actual en ciencia cognitiva, no es la única aproximación posible. Otros dos enfoques caracterizan las actividades mentales en términos de propensiones a comportarse o en términos de procesos neurales. El enfoque sobre la conducta fue característico del conductismo, que dominó gran parte de la psicología experimental (y tuvo consecuencias para la lingüística y la antropología) durante un largo período de este siglo. El enfoque conductista estaba apoyado por cierto número de argumentos filosóficos que considero en los capítulos 3 y 5. Aunque el enfoque conductista está hoy en día ampliamente pasado de moda tanto en filosofía como en psicología, el enfoque neural no lo está. En la actualidad se desarrollan serios esfuerzos para explicar la vida mental en términos de procesamiento neural. Este enfoque está apoyado también por perspectivas filosóficas que incluyen la Teoría de la Identidad Mente-Cerebro y el Materialismo Eliminativo que se discuten en el capítulo 6.

El enfoque cognitivo se caracteriza por el intento de identificar los estados mentales funcionalmente, esto es: en términos de sus interacciones causales con otros estados mentales. El reconocer la posibilidad de identificar esos estados por medio de sus interacciones causales es parte de lo que capacitó a los cognitivistas para vencer las [18-19] constricciones del conductismo. Sin embargo, la perspectiva de caracterizar esos estados independientemente de su realización material en el cerebro es lo que, para los cognitivistas, autoriza la autonomía de la psicología de la neurociencia. Durante las últimas dos décadas los filósofos han intentado desarrollar un enfoque funcionalista de los estados mentales para fundamentar el programa cognitivista. Como discuto en los capítulos 4 y 7, ha habido, sin embargo, cierto número de críticas a la coherencia de este enfoque que pueden tener, a su vez, implicaciones para el programa cognitivo.

El lenguaje ha figurado de manera central en el estudio de los procesos cognitivos. Gran parte de la teorización filosófica se ha enfocado en el lenguaje y en la capacidad del lenguaje para comportar significado. Algunos de esos puntos de vista han sido adoptados directamente en diversos programas de filosofía y lingüística, incluyendo la distinción entre el sentido de una expresión y su referente (véase el capítulo 2). Los análisis del lenguaje de la lógica formal, tales como el cálculo de predicados[4], se han empleado en los esfuerzos de la inteligencia artificial para modelar el razonamiento humano. Otros aspectos del análisis filosófico del lenguaje, tales como los desafíos a la afirmación de que las palabras tienen significados objetivos, han figurado en alguna de las críticas de la inteligencia artificial y en el desarrollo de puntos de vista recientes sobre conceptos y categorización en psicología y en lingüística.

A medida que en este texto discuto puntos de vista filosóficos distintos, señalo maneras en las que son relevantes para el trabajo en otras disciplinas de la ciencia cognitiva. Sin embargo, como indica el breve bosquejo que acaba de darse, muchos de los puntos de vista avanzados dentro de la filosofía han tenido y están teniendo ramificaciones para la ciencia cognitiva. Una consecuencia de ligar las ideas filosóficas con las investigaciones empíricas es que la evidencia empírica se torna relevante para evaluar la adecuación de puntos de vista filosóficos particulares. Esto puede sugerir, erróneamente, que la única manera que ahora existe para evaluar esos puntos de vista filosóficos es esperar los juicios de las investigaciones empíricas basados sobre ellos. Aunque ciertamente esos juicios serán relevantes, los filósofos poseen algunos recursos adicionales que pueden servir de ayuda en nuestras evaluaciones contemporáneas de esos esfuerzos. Uno de ellos es el entrenamiento para desarrollar y evaluar argumentos complejos y a menudo abstractos. Un segundo es el conocimiento [19-20] de la larga historia de los intentos de hacer frente a esos problemas. Dentro de esa historia es donde podemos a menudo localizar las fuentes de las ideas modernas. Pero, aún más importante, podemos descubrir una rica fuente de argumentos que sugieren por qué posiciones particulares son plausibles y por qué otras no son viables.

Muchas de las ideas que subyacen a los esfuerzos de investigación de la ciencia cognitiva contemporánea son descendientes directas de ideas que fueron desarrolladas anteriormente por filósofos tan antiguos como Platón, Descartes, Hume y Kant. Además, la teorización filosófica contemporánea es también la heredera de esta tradición. Por consiguiente, el resto de este capítulo ofrece una breve visión general de figuras relevantes de la historia de la filosofía, haciendo hincapié en cómo entendieron la mente y en las ideas con las que contribuyeron a las discusiones actuales.

1.2 ASPECTOS RELEVANTES DE LOS PRINCIPALES ENFOQUES HISTÓRICOS DE LA FILOSOFÍA

En una discusión breve no es posible hacer completamente justicia a ninguna de las figuras históricas más importantes de la filosofía que han tenido influencia en el pensamiento contemporáneo sobre la mente. Para presentar una explicación manejable de este material me voy a concentrar en un cierto número de tradiciones dentro de la historia de la filosofía, cada una de las cuales han ofrecido una perspectiva general sobre problemas importantes que son relevantes para nuestra comprensión de la mente. Indico brevemente algunos de los miembros más importantes de esas tradiciones y las afirmaciones centrales avanzadas por los miembros de la escuela. El lector, sin embargo, debe darse por avisado de que hay un intenso debate que rodea la interpretación de la mayor parte de esos filósofos y se necesitaría entrar en un examen detallado de esos debates para alcanzar una interpretación definitiva de cualquiera de ellos.

1.2.1 LOS FILÓSOFOS CLÁSICOS: SÓCRATES, PLATÓN Y ARISTÓTELES

Tres filósofos griegos que florecieron en los siglos V y IV a.C. establecieron la minuta de la mayor parte del pensamiento subsiguiente sobre ciencia lo mismo que sobre filosofía en el mundo occidental, incluyendo nuestros intentos de entender la mente. Sócrates planteó [20-21] las preguntas, Platón fue su

discípulo y, a su vez, el maestro de Aristóteles, pero Platón y Aristóteles ofrecieron clases diferentes de respuestas a las preguntas de Sócrates.

Sócrates (c. 470-399 a.C.) es considerado a menudo como el primero de los pensadores filosóficos importantes. Es un tanto atípico por el hecho de que no defendió ninguna tesis filosófica. Tampoco dejó escrito alguno, de modo que lo que sabemos de Sócrates surge en gran medida de la presentación que de él hizo Platón como figura central en un gran número de diálogos. Más bien que defender tesis, Sócrates desarrolló un modo de investigación al que comúnmente se hace referencia como el método socrático. Este método incluye un diálogo que comienza con una petición de definición, tal como: ¿qué es conocimiento?, o ¿qué es la belleza? Una vez que se propone la definición (p. ej., conocimiento es creencia verdadera), el que pregunta sigue planteando preguntas adicionales para evaluar la adecuación de la respuesta. A menudo esta actividad de preguntar genera contraejemplos que muestran que la definición inicial es inadecuada. (Por ejemplo: una creencia verdadera adquirida completamente por casualidad no parece ser un caso de conocimiento.) Una vez que se encuentra que la definición es deficiente, el que pregunta pide una nueva definición que supere las objeciones del intento previo, y el proceso se repite. Para Sócrates, la meta de esta actividad era descubrir definiciones universalmente verdaderas para nuestros conceptos: Al buscar tales definiciones Sócrates se oponía a los sofistas, muchos de los cuales mantenían que las definiciones precisas eran imposibles puesto que las palabras significaban cosas diferentes en contextos diferentes.

Sócrates se centró en intentar definir términos éticos como *virtud* y *justicia*, pero el método puede aplicarse claramente a cualquier concepto. Sócrates mantenía que no podemos adquirir conocimiento en ningún campo hasta que no desarrollemos definiciones adecuadas de los conceptos empleados en ese campo. El problema de si hay definiciones de nuestros conceptos que cumplan los requisitos de adecuación de Sócrates es claramente un problema crítico para la ciencia cognitiva. Los primeros científicos cognitivos, especialmente en inteligencia artificial, tendían a suponer que había tales definiciones y que podrían codificarse en programas. Además, muchos estudios del significado .y de la semántica hechos por filósofos y lingüistas han supuesto que nuestros conceptos pueden definirse. Pero los desarrollos recientes en psicología (Rosch, 1975) y en lingüística (Lakoff, 1987), así como en filosofía (Wittgenstein,Í953), han desafiado el punto de vista de que la mayor parte de nuestros conceptos está fundada en el género de definiciones que Sócrates buscaba.[21-22]

Sócrates jamás pareció encontrar definiciones adecuadas[5], pero la búsqueda fue reasumida por Platón (c. 428-347 a.C.), que pensaba que podría proporcionar una armazón para responder a las preguntas de Sócrates. Una de las objeciones frecuentes de Sócrates era que, al intentar proporcionar definiciones, los interlocutores solían citar ejemplos. Encontraba que las definiciones eran tan inadecuadas como los ejemplos puesto que no nos decían el rango de cosas al que el concepto habría de aplicarse. Así, un ejemplo de una acción justa no nos decía qué otras acciones eran justas. Platón vio que la exigencia de Sócrates de definiciones generales carecía de respuesta en la medida en que nos limitábamos al mundo físico. Propuso, por tanto, la existencia de un mundo abstracto de Ideas o Formas. Esas entidades proporcionarían los ejemplares perfectos para nuestros conceptos, y podríamos juzgar sus instancias en este mundo como imitaciones más o menos buenas de esas Ideas. Entonces, según Platón, para responder a la petición socrática de una definición, era necesario identificar la Idea, no una de sus instancias mundanas. La condición humana es, sin embargo, tal que todo lo que experimentamos son

ejemplos imperfectos de los conceptos que encontramos en el mundo físico que está a nuestro alrededor. Jamás vemos una verdadera línea recta, sino solamente una aproximación imperfecta a una línea recta trazada sobre un papel. Para clarificar nuestro pensamiento, mantenía Platón, necesitamos volver a dirigir nuestro pensamiento hacia las Ideas mismas y no permanecer concentrados en los objetos del mundo físico.

Para explicar cómo nuestro conocimiento está basado en las Ideas, Platón desarrolla una elaborada explicación de cómo alguna vez percibimos las Ideas directamente, pero debido al nacimiento hemos olvidado esta experiencia. Es necesario volver a encender esos recuerdos de modo que podamos fundamentar nuestro pensamiento en las Ideas mismas. Los objetos físicos de la experiencia, dado que son imitaciones de las Ideas, pueden facilitar este nuevo encendido si llevamos a cabo el género correcto de investigación socrática sobre esos objetos y no nos preocupamos de las distorsiones que esas imitaciones inducen. En el diálogo Menón, Platón intenta mostrar cómo el conocimiento de los principios matemáticos es innato en un esclavo no instruido, pero ha de ser sonsacado mediante una investigación en la que el esclavo comprueba la adecuación de varias [22-23] hipótesis que él mismo avanza hasta que es capaz de reconocer de una vez los principios verdaderos incorporados en las Ideas. (Para los diálogos de Platón ver Hamilton y Cairns, 1961.)

La teoría de las Ideas de Platón y su propuesta de que el conocimiento de esas Ideas es innato ha constituido un legado permanente tanto en filosofía como en otras disciplinas de la ciencia cognitiva en el contexto de la teorización sobre el conocimiento innato. La propuesta de que cierto conocimiento es innato se plantea generalmente cuando parece imposible explicar cómo podríamos adquirir ese conocimiento mediante la experiencia. Chomsky (1959) argumentó que el conocimiento de las reglas sintácticas tenía que ser innato puesto que un niño no tiene experiencia suficiente para aprenderlas por inducción. Similarmente, Fodor (1975) ha argumentado que los conceptos tienen que ser innatos puesto que no hay un modo concebible de cómo podemos aprenderlos. (Véase además Fodor, 1981; Stich, 1979; y artículos en Piatelh-Palmarini, 1980.)

Una de las afirmaciones más controvertidas de Platón es que nuestro conocimiento es realmente sobre Ideas abstractas, no sobre cosas de este mundo. Esta afirmación ha tenido su impacto más duradero en las ciencias altamente teóricas, particularmente en matemáticas. En geometría no es algo inusual el pensar en figuras puras como triángulos existentes de manera separada de cualquier dibujo de ellos. Del mismo modo, la distinción entre números y numerales parece capturar la distinción entre los objetos puros y nuestras representaciones de ellos. Pero muchos han encontrado insostenible la conclusión platónica de que nuestro conocimiento no es de cosas de este mundo. Platón mismo presentó alguna de esas dificultades en sus últimos diálogos, pero fue su pupilo Aristóteles (384-322 a.C.) el que se encargó de subrayarlas y ofreció un esquema filosófico alternativo que volvió a dirigir la atención hacia los objetos de este mundo. Aristóteles conservó algo de la noción platónica de las Ideas con su concepto de las Formas, pero argumentó que las Formas están en los objetos de los que tenemos experiencia, no en algún espacio abstracto. Aristóteles interpretó los objetos del mundo como consistiendo de una Forma impuesta sobre una materia (p. ej., una taza consta de la imposición de la Forma TAZA sobre el barro de la que está hecha). Mantuvo que la Forma determinaba el género de objeto que algo era y fijaba muchas de sus propiedades básicas.

El adquirir conocimiento de un objeto exigía para Aristóteles el reconocimiento de la Forma que

había en él. Al igual que los científicos cognitivos modernos, Aristóteles se interesaba por cómo podemos representar en nuestras mentes los objetos del mundo. Desarrolló una teoría de la percepción por medio de la cual la Forma que [23-24] definía el objeto se transfería a la mente del que percibe. Así pues, percibir una mesa requería captar efectivamente la forma del objeto (pero no su materia) en la mente del que percibe. Aristóteles mantuvo entonces una versión primitiva de la teoría representacional (ver capítulo 4).

La explicación aristotélica de las Formas fue crucial para las teorías científicas que él desarrolló y que se mantuvieron hasta el siglo XVII . Él permitía que la Forma que definía a un objeto pudiese cambiarse como, por ejemplo, cuando un carpintero toma un árbol y hace de él una mesa. Por otra parte, las Formas son las que proporcionan la organización y el principio rector de los objetos naturales de modo que esos objetos se comportan de acuerdo con sus formas. La Forma del objeto, al menos en el caso de los organismos vivientes, especificaba la meta hacia la que se estaba desarrollando (ver capítulo 5). A este respecto el punto de vista de Aristóteles sobre la naturaleza es completamente diferente del moderno. Para Aristóteles (lo mismo que para Platón[6]), la naturaleza es ideológica o dirigida hacia un fin. Mientras que nosotros vemos generalmente los objetos como pasivos, él los contemplaba como buscando ciertos objetivos determinados por su Forma. Cuando Aristóteles intentaba analizar el cambio en la naturaleza, no se concentraba en lo que nosotros llamaríamos la «causa» del cambio, sino en cuatro factores: la materia que subyacía al cambio, el evento que indujo el cambio, la forma que se hacía real como resultado del cambio, y la meta hacia la que se dirigía el cambio[7]. Las aplicaciones de este punto de vista se encuentran en la explicación aristotélica de cómo diferentes géneros de objetos buscan su propio lugar en la naturaleza (p. ej., el fuego lucha por subir hacia arriba mientras que la tierra tiende a ir hacia el centro del universo) y en su concepción de las cosas vivientes como algo que busca dar cumplimiento a sus formas. (Para los escritos de Aristóteles, ver McKeon, 1941.)

La ciencia moderna, que se ha desarrollado desde el siglo XVII, ha repudiado la idea de un universo orientado teleológicamente a favor de un modelo mecanicista. Aunque ha resultado bastante fácil eliminar la noción de teleología de nuestras explicaciones de los fenómenos puramente físicos, ha sido mucho más difícil en las explicaciones de los fenómenos biológicos y cognitivos, pues éstos parecen ser fenómenos dirigidos hacia una meta. Así pues, uno de los problemas filosóficos que tenemos que afrontar al dar un análisis conceptual adecuado de la biología moderna y de la ciencia cognitiva es proporcionar una armazón que pueda acomodar el carácter teleológico de las cosas vivientes y de los sistemas cognitivos sin ir más allá del tipo de armazón mecanicista desarrollada originalmente dentro de las ciencias físicas (ver capítulo 7).

Los puntos de vista de Sócrates, Platón y Aristóteles, aunque ya no se acepten en su forma original, continúan teniendo influencia en el pensamiento de la ciencia cognitiva en una gran variedad de maneras. Además han tenido un impacto duradero sobre la ciencia e incluso un impacto mucho más amplio sobre nuestra ciencia popular (McCloskey, 1983). La explicación aristotélica de los objetos proporcionaba en particular una estructura comprensiva en la que describir y categorizar los fenómenos naturales que sirvió como base para la ciencia hasta el siglo XVIl. Lo que, sin embargo, no proporcionaba era una estructura adecuada para entender el proceso dinámico de la naturaleza. La revolución científica incluyó en gran medida el desarrollo de un punto de vista dinámico de la naturaleza en el que el punto focal no era la identificación de la esencia de los objetos, sino el modelar el cambio en términos de los movimientos

inducidos en la materia física. Esto incluía el desarrollo de una concepción mecánica del universo. Se desarrollaron dos nuevas perspectivas filosóficas —*Racionalismo* y *Empirismo*— como intentos de proporcionar una armazón conceptual para la comprensión de la nueva ciencia mecanicista de Copérnico, Galileo y Newton. Aunque la mente no se consideraba como un objeto central de estudio científico en esta nueva ciencia, las explicaciones racionalistas y empiristas de cómo podríamos conocer las afirmaciones de esta ciencia han tenido un impacto duradero en la teorización sobre la mente.

1.2.2. RACIONALISMO

El racionalismo surgió como la tradición filosófica dominante en Europa durante los siglos XVII y XVIII. Sus tres más famosos representantes fueron Descartes (1596-1650), Leibniz (1646-1716) y Spinoza (1632-1677). Para entender a los racionalistas debemos tener presente que estuvieron profundamente implicados tanto en el desarrollo efectivo de la ciencia moderna, como en proporcionar una explicación filosófica coherente de ella. Hoy día sus puntos de vista filosóficos se toman en consideración independientemente de sus [25-26] contribuciones al desarrollo de la ciencia, pero esto representa de manera desacertada su enfoque de la filosofía.

Lo que distingue al racionalismo es una profunda confianza en la razón como instrumento para descubrir los procesos que operan en la naturaleza. Para los racionalistas los sentidos tienen algún papel que desempeñar, pero éste es secundario respecto al de la razón. Parte de la atracción que la razón ejercía sobre los racionalistas se debía a su convicción de que la naturaleza tenía que haber sido diseñada de una manera lógicamente inteligente. Si esto era verdad, entonces una investigación lógica cuidadosa podría llevarnos a las verdades fundamentales. El carácter de tal investigación lógica está ejemplificado en las *Meditaciones* (1641/1970) de Descartes. Comienzan las Meditaciones con un programa de duda radical mediante el cual Descartes cuestionaba toda creencia de la que no estuviese seguro. Para extender esta duda máximamente, Descartes contemplaba la posibilidad de que estuviese bajo el control de un genio maligno cuyos esfuerzos estuviesen dirigidos a engañarlo lo más posible. Descartes afirma que el motivo de suscitar esas dudas era limpiar su mente de todas las proposiciones dudosas que no hubiesen sido demostradas completamente. Él atribuyó muchos de sus pensamientos erróneos sobre la naturaleza a la aceptación sin cuidado de ideas que no habían sido cuidadosamente examinadas.

Una vez que el terreno se había limpiado de ideas erróneas, la meta de Descartes era construir un nuevo edificio de verdades científicas que se razonaría cuidadosamente a partir de cimientos indubitables. La primera verdad indubitable que él pensó que había descubierto era su propia existencia, que consideró que era una consecuencia necesaria del hecho de que estaba pensando cuando planteaba esas dudas. Incluso el genio maligno no podría amañar una situación en la que Descartes pensase algo y a la vez no existiese. (Éste es el contexto de la famosa expresión de Descartes «*Cogito ergo sum*» o «Pienso, luego existo»).

Al establecer que de su existencia no se podía dudar, Descartes pensó que había descubierto un método para establecer afirmaciones sobre aquello de lo que podía estar seguro. Afirmó que la idea de su existencia era «clara y distinta». Para él una idea era clara cuando captábamos su esencia; era distinta cuando la percibíamos diferenciada de otras ideas. Descartes formó la hipótesis de que todas las ideas

claras y distintas son verdaderas y se marcó la tarea de justificarlo. Para hacer esto intentó mostrar que la idea de un genio maligno era incoherente y que en su lugar había un Dios benevolente que le proporcionaba sus ideas. Una vez que hubo logrado esto, razonó que, puesto que Dios era benevolente, podía confiar en sus ideas en [26-27] la medida en que se adhiriese a los principios del razonamiento apropiado al sintetizar conocimiento a partir de sus ideas claras y distintas. Así pues, el método de razonar mediante ideas claras y distintas quedaba vindicado.

El argumento de Descartes a favor de la existencia de Dios ha sido muy criticado en la literatura filosófica, pero esto no nos obliga a distraernos de su programa general, que era desarrollar los fundamentos conceptuales de la nueva física. Lo que Descartes pensaba que mostraban sus ideas claras y distintas era que la naturaleza era un sistema corpuscular (ver Descartes, 1644/1970). Todos los objetos físicos estaban compuestos de finos corpúsculos, y las propiedades básicas de esos corpúsculos —sus tamaños, formas y movimientos— determinaban la conducta de los objetos físicos. El movimiento de un corpúsculo resultaba de las fuerzas que incidían sobre él a partir de colisiones con otros corpúsculos. Además, Descartes razonaba que no podía haber espacio que no estuviese ocupado por corpúsculos y que todas las interacciones entre corpúsculos resultaban del contacto físico directo. En términos de estos principios básicos, Descartes intentó desarrollar teorías que podrían explicar la conducta observada de los objetos físicos. Pensó que casi todos los fenómenos naturales, animados e inanimados, podrían explicarse así: en términos de interacciones físicas de corpúsculos. Descartes hizo una excepción solamente: el caso de la mente humana (ver capítulo 5). Ésta fue la fuente del «dualismo» cartesiano (el punto de vista de que la mente está separada del cuerpo), pero, desde su punto estratégico de intentar proporcionar un fundamento para la física que pudiese explicar la naturaleza, esto era una excepción relativamente menor.

Me he centrado en Descartes porque su programa es un prototipo de las preocupaciones de los racionalistas. Desde el punto de vista estratégico de la ciencia cognitiva, lo más importante del programa racionalista no es el intento de proporcionar certeza a nuestro conocimiento, sino el énfasis en la importancia del razonamiento para llegar a nuestro conocimiento. Los racionalistas, al igual que Platón con anterioridad, tomaron su modelo de conocimiento de los matemáticos, que intentaban derivar teoremas de principios que consideraban indubitables. Aunque la suposición de que los postulados matemáticos son indubitables ha sido desafiada durante los dos pasados siglos, la concepción de las matemáticas como algo que descansa en razonamiento lógico a partir de postulados ha permanecido. Muchos científicos cognitivos han compartido el punto de vista de que la cognición es primariamente un proceso de razonamiento. Esto es particularmente verdadero de los que se ocupan de la inteligencia artificial (IA) que han diseñado programas en los que se codifican [27-28] principios básicos del conocimiento y se extraen conclusiones mediante diversos recursos del razonamiento lógico. Aunque el aspecto materialista de la afirmación de que un computador puede simular el razonamiento era extraño a Descartes, la capacidad del computador de llevar a cabo inferencias lógicas recomendaría al racionalista el usar el computador como un instrumento para modelar el pensamiento. Del mismo modo, no es sorprendente que un lingüista como Chomsky (1966), que piensa que las estructuras del lenguaje se producen mediante la aplicación de reglas, caracterice su programa como «lingüística cartesiana».

1.2.3. EMPIRISMO

Mientras que el racionalismo se estaba desarrollando en el continente europeo, un punto de vista radicalmente diferente, conocido como empirismo, se desarrolló en las islas Británicas durante los siglos XVII y XVIII . Aunque es aún importante, la razón desempeña un papel bastante menos central para los empiristas. La percepción sensorial proporciona, en su lugar, el fundamento. Un precursor del movimiento empirista, Francis Bacon (1561-1626), atribuyó los errores de la ciencia aristotélica a un exceso de confianza en la razón, y argumentó que sólo mediante una total y absoluta fidelidad a la evidencia sensorial se podría fundamentar el edificio de la nueva ciencia. El propósito de Bacon era construir conocimiento de verdades generales siguiendo los principios de la inducción tomando como base la evidencia proporcionada por los sentidos (Bacon, 1620).

En muchos aspectos Locke (1623-1704) estableció el modelo de análisis para los empiristas. Afirmó que todo el conocimiento se remontaba a la experiencia sensorial e intentó mostrar cómo la experiencia da lugar a ideas simples o elementales. Expuso también cómo la mente asocia ideas de objetos particulares para formar ideas complejas, así como también ideas generales y abstractas, necesarias para la ciencia. El principio de que la mente opera principalmente asociando ideas simples a partir de la experiencia proporcionó la base para una duradera tradición que los científicos cognitivos reconocen como *Asociacionismo* (Locke, 1690/1959).

De entre los empiristas más importantes, Locke era el mayor devoto de la ciencia newtoniana. Su objetivo era mostrar cómo se podía fundamentar la ciencia de Newton en una epistemología empirista que comenzase con la experiencia y desarrollase todo el conocimiento restante mediante principios de asociación. En particular Locke pensaba que podría justificar el punto de vista newtoniano básico de [28-29] un universo mecanicista que operase de una manera muy semejante a la de un reloj[8]. En contraste, tanto Berkeley como Hume desafiaron algunas de las características de la ciencia newtoniana e intentaron colocarla bajo lo que ellos creían que era una mejor luz.

Berkeley (1685-1753) estaba espantado por la posibilidad de que el punto de vista mecanicista de Newton no dejase lugar alguno para Dios[9]. Su remedio a la mecanización del mundo era radical: negó la existencia del mundo físico como un objeto existente fuera del pensamiento. Argumentó que la afirmación de que nuestras ideas son sobre objetos físicos externos a nuestras ideas era incoherente, manteniendo que nuestros pensamientos jamás nos podrían informar sobre algo excepto sobre nuestras ideas. Así pues, jamás podríamos saber nada sobre un mundo físico que exista separadamente, si tal mundo existiese. Además, la verdad de la ciencia, argumentaba Berkeley, no depende de la existencia de un mundo físico externo. Las ideas y las mentes que las piensan eran entonces los únicos objetos que se necesitaban. Berkeley apelaba a Dios para explicar las regularidades y la coherencia entre las ideas que adquirimos a partir de la experiencia sensorial. Incluso aunque no tengamos ideas de las cosas, Dios podría tenerlas y los objetos podrían existir, por tanto, en la mente de Dios. Entonces, aunque niegue la existencia de un mundo externo, físico, Berkeley no niega la existencia de objetos y la legitimidad de las investigaciones científicas. Simplemente mantenía que esos objetos estaban presentes en ideas y que aquello sobre lo que la ciencia versaba era el orden de las ideas tal como nos eran presentadas por Dios (Berkeley, 1710/1965).

Hume (1711-1776) se separaba del esquema newtoniano en una dirección diferente. Al igual que Descartes, Hume comienza su investigación en vena escéptica. Desafiaba nuestras afirmaciones de que

conocemos una serie de cosas que mucha gente afirma que conoce. Uno de sus blancos principales era la *causalidad*. Hume argumentaba que la experiencia jamás podría revelamos las relaciones que se [29-30] dan entre causa y efecto. La experiencia puede mostrarnos que un tipo de evento es seguido regularmente por otro, pero no que hay conexión intrínseca alguna entre ellos. Al hacer esta afirmación, Hume estaba minando un principio fundamental de la nueva ciencia newtoniana, pero argumentó que las consecuencias no eran tan drásticas como podría parecer. Al ser incapaz de encontrar fundamentos experimentales de ningún tipo para nuestra creencia en la causalidad, Hume la retrotrajo a una disposición natural en los seres humanos para formar asociaciones entre eventos que aparecen regularmente unidos en la experiencia. Nuestras creencias sobre las relaciones causales no son algo sobre el mundo que pueda inferirse razonando sobre nuestra experiencia sensorial, sino que son simplemente reflejos de nuestro carácter básico y del modo en que experimentamos la naturaleza (Hume, 1748/1962, 1759/1888; para una discusión de las contribuciones de Hume a la ciencia cognitiva, ver Biro, 1985a).

Aunque alcanzaron esta conclusión de maneras diferentes, tanto Berkeley como Hume mantuvieron que ajustarse al principio empirista básico de retrotraer todas las afirmaciones de conocimiento a las experiencias sensoriales y a las inferencias que extraemos de ellas daba como resultado mayores restricciones que las que Locke pensaba sobre lo que podía conocerse. En esto ellos se vieron a sí mismos como empiristas mucho más cabales que Locke. El imponer límites sobre lo que los humanos pueden conocer ha sido parte del legado más permanente de los empiristas. Vemos esto tanto en el asociacionismo como en el conductismo, que, en tanto que herederos del empirismo han argumentado a favor de establecer límites sobre lo que podemos conocer basándose en teorías acerca de cómo adquirimos conocimiento a partir de la experiencia.

1.2.4. EL PUNTO DE VISTA KANTIANO

De todas las figuras históricas de la filosofía es Kant (1724-1804) el que ofreció puntos de vista que se alinean más estrechamente con los avanzados por la ciencia cognitiva, si bien él no habría dado su aprobación a ésta. Kant puede verse en parte como una síntesis de las tradiciones empirista y racionalista. Comenzó intentando responder a Hume. Veía que el escepticismo de Hume llevaba a resultados desastrosos, particularmente porque minaba la potencialidad de conocer las relaciones causales de la naturaleza postuladas por la ciencia newtoniana. Consideró nuestra capacidad de conocer la ciencia newtoniana como algo dado y se marcó la tarea de mostrar cómo tal conocimiento era posible. Estaba de acuerdo con Hume y otros empiristas en que nuestro conocimiento de los procesos físicos depende de [30-31] la experiencia y en que no se descubre simplemente razonando sobre nuestras ideas innatas. Sin embargo, vio también que el escepticismo de Hume era la consecuencia inevitable de la adhesión al principio empirista que intentaba extraer todo él conocimiento de la experiencia. La única solución que vio fue lanzar su «Revolución copernicana» en filosofía mediante la que dio la vuelta a la relación de los humanos con el mundo natural. Mientras que toda la filosofía anterior suponía que los objetos de conocimiento existen independientemente de nosotros y, a continuación, preguntaba cómo podíamos conocerlos, Kant mantuvo que nuestras actividades cognitivas eran parcialmente constitutivas de los objetos de los que tenemos experiencia. Mantiene además que es precisamente nuestra propia participación en la construcción de los objetos de percepción lo que hace posible que los conozcamos.

Al explicar cómo nuestra actividad cognitiva es constitutiva de los fenómenos que experimentamos, Kant suscribió en parte el enfoque racionalista. Afirmaba que nuestra capacidad de percibir y de pensar sobre la naturaleza dependía de conceptos o categorías del entendimiento que aportamos a la experiencia, categorías que poseemos de manera innata. Pero las categorías que Kant tenía en mente no eran las categorías mediante las que clasificamos objetos. Más bien, sus categorías especifican el carácter general de los objetos y las relaciones en las que están. Así, él incluye *causa* y *efecto* como una categoría. Además esas categorías no están representadas en la mente como conceptos que puedan analizarse para derivar conocimiento de la naturaleza, tal como el racionalismo mantenía. Más bien, esas categorías tenían que aplicarse al *input* sensorial que recibimos para constituir un mundo de experiencia. Para hacer esto posible, Kant mantenía que las categorías tenían que esquematizarse, esto es: necesitaban que se les diese interpretaciones en términos del carácter espacio-temporal necesariamente exhibido por todos los estímulos sensoriales. El esquema para la causa es, por ejemplo, la sucesión constante de un estado por otro. Para que nosotros tengamos experiencia de un objeto, el intelecto tiene que aplicar las categorías esquematizadas a nuestro *input* sensorial. Así, los objetos que experimentamos son el producto de aplicar las categorías esquematizadas a *inputs* sensoriales brutos. Nuestro conocimiento se limita a esos objetos construidos.

Kant mantuvo que la experiencia sensorial bruta que no se somete bajo las categorías y los objetos que dan lugar a esas experiencias sensoriales (que Kant denominó cosas en sí) es incognoscible para nosotros. Por consiguiente, no tiene sentido investigar qué son realmente las cosas en sí. Por otra parte, los objetos de la experiencia [31-32] fenoménica, los que se construyen aplicando las categorías a los estímulos sensoriales, están dentro de nuestro dominio de conocimiento. Puesto que esos objetos se han construido de acuerdo con nuestras categorías, podemos estar seguros de que se adhieren a los principios establecidos en esas categorías. Por ejemplo, puesto que construimos el mundo de modo que cada evento tenga una causa, sabemos con certeza que todo evento tiene una causa. Puesto que principios como el de causación se usan al construir el mundo, Kant afirmaba que podíamos saber con certeza que los principios de la física de Newton son verdaderos.

Kant llamó a los principios que son el resultado necesario de aplicar las categorías a la experiencia *sintéticos a priori*. Para explicar lo que quiere decir mediante esto, servirá de ayuda el colocar en perspectiva la posición de Kant y mostrar cómo está ligada a la ciencia cognitiva moderna. Previamente había distinguido entre el conocimiento *a priori*, lo que es cognoscible sin la experiencia, del conocimiento *a posteriori*, que depende de la experiencia. Necesitamos ahora introducir una segunda distinción entre enunciados analíticos y sintéticos. Los enunciados *analíticos* son enunciados que son verdaderos en virtud del significado de las palabras. Por ejemplo, el enunciado «un soltero no está casado» es verdadero en virtud del significado de la palabra «soltero». Los enunciados *sintéticos* son aquellos que juntan conceptos de manera que pueden ser falsos. Por ejemplo, el enunciado «el coche es rojo» no es verdadero en virtud del significado y puede ser falso. Sólo los enunciados sintéticos hacen afirmaciones sustantivas sobre el mundo.

Es tradicional pensar que los enunciados analíticos se conocen *a priori* puesto que dependen del significado de las palabras, y que los sintéticos se conocen *a posteriori* porque hacen afirmaciones sustantivas sobre el mundo y así exigen de la experiencia para ser conocidos. Kant rechazó este punto de vista y trató algunos enunciados sintéticos como cognoscibles *a priori*. Mantiene, pues, que antes de la

experiencia efectiva podemos conocer cómo tienen que ser las cosas en la naturaleza. Esto se debe al papel que desempeñan las categorías en el mundo en que experimentamos objetos. En el vocabulario de la ciencia cognitiva moderna, Kant está introduciendo procesamiento de arriba-abajo en nuestros procesos cognitivos, incluyendo la percepción, y está afirmando que este procesamiento está constriñendo el proceso de conocimiento. Kant, sin embargo, no reconocería como suya muy probablemente esta interpretación de su punto de vista puesto que las concepciones sobre el procesamiento de la ciencia cognitiva moderna se considera que son partes de la ciencia empírica, mientras que él pensaba que el papel de las categorías en la cognición [32-33] no podía estudiarse empíricamente, sino sólo averiguarse investigando las condiciones necesarias para la experiencia. (Véase, sin embargo, Biro, 1985b.) Kant habló de tal investigación como trascendental (Kant, 1787/1961).

El propósito de Kant constituía una línea divisoria en el pensamiento filosófico puesto que abría la posibilidad de que el mundo que conocemos sea el mundo que construimos y no algún mundo independiente con el que hemos de luchar para entrar en contacto. Uno de los puntos de vista de Kant que resultó más controvertido fue su afirmación de que los conceptos y categorías que identificaba eran aquellos que tenían qué usarse para tener cualquier experiencia. Así pues, él pensaba que no sólo la ciencia newtoniana sino también la geometría euclídea eran necesariamente verdaderas, no sólo empíricamente verdaderas. La introducción de geometrías no euclídeas y, de manera subsiguiente, la de la física no newtoniana arruinó la suposición de que las categorías de Kant eran necesarias.

Entre las diversas modificaciones del enfoque de Kant que han sido consideradas, una de las más influyentes fue el desarrollo del *pragmatismo*, particularmente por medio de la obra del filósofo norteamericano Charles Sanders Peirce (1839-1914). Peirce renunció a la pretensión de que hay un conjunto de categorías que tenemos que emplear para conceptualizar la naturaleza, pero mantuvo con Kant que nosotros proporcionamos de hecho los conceptos organizadores que usamos para conceptualizar la naturaleza. En lugar de argumentar que esos conceptos están legitimados porque son aquellos que tenemos que usar, Peirce propuso que éstos ganan legitimidad en la medida en que prueban ser fructíferos en nuestro intento de desarropar teorías adecuadas de la naturaleza. Peirce se concentra en la investigación como un proceso con actividad correctiva. Para Peirce, los investigadores adoptan conceptos y teorías y tratan de organizar sus experiencias en términos de ellos. Esos conceptos y teorías dan lugar a expectativas, expectativas que pueden fallar. Cuando fallan, los investigadores tienen que modificar sus conceptos y teorías para generar expectativas que estén en mejor acuerdo con lo que sucede. Es ésta una empresa activa, pero es una empresa que, afirma Peirce, proporcionará finalmente un conjunto de conceptos y teorías que no exigirán una modificación subsiguiente. Aunque no sabremos cuándo hemos alcanzado el punto en el que ninguna experiencia futura contravendrá nuestras expectativas, cuando lo alcancemos tendremos conocimiento de cómo es el mundo[10]. (Véase Peirce. 1877/ 1934, 1878/1934.) [33-34]

1.2.5. DOS TRADICIONES CONTEMPORÁNEAS: CONTINENTAL Y ANALÍTICA

Al igual que durante muchos otros períodos de la historia, la comunidad filosófica del mundo occidental está dividida actualmente en dos enfoques diferentes. La tradición *analítica* ha sido la

tradición más importante en el mundo de habla inglesa durante este siglo, y ha atraído periódicamente a filósofos en Alemania, Holanda y Escandinavia. Por contraste, la tradición *continental* ha sido muy influyente en Europa, aunque ha atraído cada vez más interés en el mundo de habla inglesa.

La mayor parte de la obra sobre filosofía de la mente que ha sido discutida por los científicos cognitivos se ha originado dentro de la tradición analítica. Los puntos de vista filosóficos que se describen en los capítulos que siguen proporcionan, por tanto, una introducción al carácter de la filosofía analítica. Aquí hago observar simplemente dos de los factores que han dado forma al desarrollo de esta tradición.

Uno es la confianza en el uso de la lógica simbólica como instrumento para el análisis. (Véase Bechtel, en prensa b, para una breve introducción a la lógica simbólica y a la manera en cómo figura en la filosofía de la ciencia moderna.) Otro es un interés por el lenguaje. Este interés ha tomado dos formas. Por una parte, los filósofos analíticos han pensado a menudo que los problemas filosóficos podrían resolverse clarificando nuestro uso del lenguaje. Como resultado de esto los filósofos analíticos se han entregado a menudo a la práctica del análisis conceptual, intentando clarificar el significado de conceptos particulares tales como *creencia*, *libertad* o *verdad*. Por otra parte, los filósofos analíticos han estado interesados en el lenguaje mismo y han buscado dar cuenta de cómo funciona. En particular, los filósofos analiticos se han interesado en cómo las palabras tienen significado de modo que las oraciones puedan decir cosas. En el próximo capítulo se describen varias explicaciones del lenguaje avanzadas por los filósofos analíticos.

La tradición continental ha estado menos comprometida con el análisis lógico del lenguaje y mucho más interesada en la descripción exacta de los rasgos básicos de la existencia humana. Dentro de la tradición continental ha habido dos escuelas centrales que se han enfocado hacia aspectos diferentes de la experiencia humana. La escuela *fenomenológica* surgió a fines del siglo XIX por medio de la obra de filósofos como Hüsserl, y ha continuado a través de filósofos como [34-35] Merleau-Ponty. Ha buscado analizar el contenido de la experiencia humana y los procesos mediante los que nuestras experiencias fenoménicas toman forma. La escuela *existencialista*, representada por filósofos como Heidegger y Sartre, se ha enfocado más hacia el contexto de la experiencia y de las exigencias para actuar en tales contextos. De este modo, Sartre habló de que los humanos se encuentran a sí mismos arrojados a la existencia con la necesidad de crear para sí mismos principios mediante los que tomar decisiones.

Más recientemente ha surgido un nuevo movimiento en la tradición continental. La escuela *hermeneútica*, asociada con Derrida, subraya el proceso de interpretación tanto de textos como de la cultura en general. La idea básica es que se debe «desconstruir» el texto o la cultura de modo que se descubran las suposiciones fundamentales que se hacen en ella. Esas suposiciones no han de ser justificadas o refutadas, sino simplemente expuestas.

1.3 CONCLUSIÓN: PREPARADOS PARA HACER FRENTE A LOS PROBLEMAS

Este capítulo ha constituido una preparación para el propósito principal de este libro: proporcionar una introducción a la filosofía de la mente contemporánea. He caracterizado brevemente los propósitos de la filosofía de la mente con respecto tanto al método filosófico de plantear problemas sobre la mente, como a la relevancia de los puntos de vista filosóficos para la misma ciencia cognitiva. Proporciono

también breves explicaciones de las figuras más importantes de la historia de la filosofía que son relevantes para la teorización y la investigación filosófica actual en la ciencia cognitiva. En el capítulo siguiente discuto la investigación en filosofía del lenguaje que ha contribuido a la filosofía de la mente y ha tenido influencia en varias de las ciencias cognitivas, incluyendo la lingüística y la inteligencia artificial. [35-36]

2. ANÁLISIS FILOSÓFICOS DEL LENGUAJE

2.1 INTRODUCCIÓN

El análisis del lenguaje ha sido la preocupación más importante de los filósofos analíticos. Sin embargo, los filósofos no han sido los únicos investigadores que han intentado analizar el lenguaje y así, con el objeto de establecer la armazón para discutir la filosofía del lenguaje, es útil indicar cómo difieren los análisis filosóficos del lenguaje de los avanzados por otras disciplinas de la ciencia cognitiva. Los psicólogos se han interesado principalmente en los procesos internos a la mente que hacen posible el uso del lenguaje. En contraste, los filósofos han contemplado el lenguaje como un objeto digno de ser analizado por sí mismo, sin plantear cuestiones sobre procesos psicológicos internos. A este respecto la filosofía del lenguaje está más próxima a la lingüística. Pero los análisis filosóficos difieren también de los de la lingüística. Los lingüistas han estado interesados principalmente en el desarrollo de caracterizaciones abstractas bien de la sintaxis o de la semántica de un lenguaje, y han producido a menudo explicaciones generativas que intentan predecir el conjunto infinito de oraciones que pueden surgir en un lenguaje a partir de un número finito de principios. Los filósofos, por otro lado, han intentado proporcionar explicaciones generales de lo que constituye el significado de las expresiones lingüísticas sin intentar desarrollar teorías detalladas para dar cuenta de los tipos de emisiones que aparecen en los lenguajes efectivos. Aunque las aspiraciones de los filósofos, psicólogos y lingüistas son distintas, sus propósitos están claramente relacionados de modo que las contribuciones en una disciplina han sido empleadas en las otras.

Los filósofos han desarrollado de modo efectivo una gran variedad de análisis diferentes y en competición del significado lingüístico desde el pasado siglo. Mi discusión sigue el orden histórico en el que esas ideas fueron avanzadas. En muchos casos se propusieron análisis subsiguientes para solucionar problemas, o problemas que se percibían por vez primera, en los análisis anteriores. Esto no significa que los últimos análisis sean superiores y que las primeras posiciones tengan meramente un interés histórico. Muchos filósofos aún apoyan las posiciones primitivas y han intentado superar objeciones que se [36-37] les han hecho a ellas. Por tanto, cada explicación del significado lingüístico debería evaluarse por su adecuación y no descartarse porque otros puntos de vista se hayan puesto de moda.

2.2. ANÁLISIS REFERENCIALES DEL SIGNIFICADO: MEINONG, FREGE, RUSSELL Y EL PRIMER WITTGENSTEIN

La preocupación por el significado de las palabras y las oraciones del lenguaje surgió muy al comienzo de la filosofía analítica con la obra de Meinong, Frege, Russell y Wittgenstein. Estos filósofos hicieron de la *referencia* —el fenómeno consistente en que las palabras se refieran o denoten algo en el mundo— el punto central de sus análisis del significado. El significado de una palabra como «martillo», mantenían ellos, consistía en el objeto, un martillo, al que esa palabra se refería.

Los filósofos que abogaban por el enfoque referencial fueron justamente los mismos que fueron los responsables del desarrollo de la lógica simbólica moderna. Su análisis referencial es una consecuencia natural del de la lógica que considera como paradigmático lo que se denomina *discurso extensional*. En el discurso extensional los símbolos del lenguaje están por objetos o propiedades de objetos, y las afirmaciones que se hacen en las oraciones del lenguaje se considera que caracterizan (verdadera o falsamente) a esos objetos y sus propiedades. Los lenguajes extensionales se adhieren a lo que es conocido como la *Ley de Leibniz* de acuerdo con la cual podemos sustituir un término por otro que se refiere al mismo objeto sin cambiar el valor de verdad de la oración. Por ejemplo, en la oración «el Buick verde chocó contra el Ford rojo» podemos sustituir «el Buick verde» por «el viejo coche de Lesley» si ambas expresiones se refieren al mismo coche y, si la primera oración es verdadera, entonces la segunda oración será también verdadera. En tal discurso extensional la relación de referencia entre los nombres lingüísticos de los objetos y los objetos mismos es absolutamente fundamental. Pero esta relación se torna problemática en al menos algunos casos. Los problemas se expresaron en cierto número de acertijos lógicos y las teorías del lenguaje que los primeros filósofos analíticos avanzaron fueron diseñadas para resolver esos acertijos[1]. [37-38]

Uno de los acertijos lo generó Alexius Meinong (un filósofo relacionado sólo de manera tangencial con el movimiento analítico). Su acertijo tiene que ver con juicios sobre objetos no existentes tales como el siguiente: «El cuadrado redondo no existe», o el enunciado: «La montaña de oro no existe» (Meinong, 1904/1960). La expresión «cuadrado redondo» o «montaña de oro» son los sujetos de esas oraciones, y de este modo parece que nos estamos refiriendo a un cuadrado redondo o a una montaña de oro. Esto parece paradójico porque nos estamos refiriendo al objeto y, de este modo, afirmando su existencia en el mismo acto de negar su existencia. Para resolver tales acertijos Meinong argumentó que teníamos que invocar un concepto más amplio de los objetos en el que admitiésemos objetos que no existiesen. Meinong propuso que había objetos puros, más allá del existir o no existir, que podrían ser los referentes de nuestros términos lingüísticos incluso cuando no hubiese objetos fácticos que les correspondiesen. En efecto, lo que Meinong está haciendo es distinguir la relación de referencia que se da entre un término lingüístico y su referente de las relaciones ordinarias entre objetos. Para que alguien instancie la relación ordinaria de comprar pan, debe haber efectivamente tanto el pan como la persona que lo compra. Pero esto no es verdad de la referencia: para que alguien se refiera al pan, no se exige que haya efectivamente una pieza de pan a la que se haga referencia. La solución de Meinong, que permitía que hubiese objetos distintos de los efectivamente existentes a los que nos podemos referir, sorprendió a los filósofos subsiguientes, tales como Russell y Ryle, que la consideraron peor que el problema mismo. Contemplaron a Meinong como alguien que postulaba de manera innecesaria nuevos géneros de objetos, y esto les llevó a preferir otras soluciones a este problema.

En 1892 Gottlob Frege, uno de los pensadores que contribuyeron de manera principal al desarrollo de la nueva lógica, planteó, un género distinto de problema para el análisis del lenguaje. Estaba centrado en el predicado de identidad, representado en castellano por el verbo «es» en enunciados de la forma «X es Y». Las oraciones de esta forma parecen representar una relación entre dos objetos, pero Frege muestra que las dos explicaciones más naturales de la relación de identidad no logran capturar el significado de enunciados tales como «Venus es la estrella de la mañana». Una explicación contempla [38-39] la identidad como una relación que se da entre un objeto y sí mismo. Si esto fuera el caso, entonces

podríamos sustituir igualmente un nombre por el otro, dando lugar al enunciado «Venus es Venus». Esto, sin embargo, carece de la informatividad de la oración original. La otra explicación contempla la identidad como una relación entre nombres: están en la relación de nombrar la misma cosa. Pero, de acuerdo con esta explicación, «Venus es la estrella de la mañana» no enuncia más que nuestra aceptación de una convención lingüística para usar los dos nombres de modo correferencial. Como tal no hace afirmación empírica alguna.

Para explicar tales enunciados, Frege introduce la distinción entre el *sentido* y la *referencia* de un término. El referente es el objeto nombrado o al que se hace referencia de cualquier otra manera por el término, mientras que el sentido incluye el «modo de presentación» mediante el cual se nos presenta el referente. Usando esta distinción Frege resuelve el problema sobre el enunciado «Venus es la estrella de la mañana». El enunciado nos dice que dos términos con sentidos diferentes tienen efectivamente los mismos referentes. Esto es informativo en el sentido en que «Venus es Venus» no lo es. Pero esto no enuncia simplemente una convención lingüística. Más bien lo que hace es describir un descubrimiento astronómico efectivo. Informa de que dos términos cuyos sentidos han sido fijados de antemano de modo que puedan referirse a objetos diferentes, se refieren, como se ha descubierto, al mismo objeto.

La distinción de Frege entre sentido y referencia ha sido muy influyente en análisis subsiguientes del lenguaje, de modo que resulta útil desarrollar algunos otros aspectos de su discusión. Frege propuso que en ciertos contextos un término podía cambiar de tener su sentido y referencia *habituales* a tener un sentido y referencia *indirectos*.La referencia indirecta de un término es su sentido ordinario. Esto le permite a Frege resolver otro problema lógico que surge en oraciones que contienen verbos como «sabe», «cree» y «piensa» seguidos de una proposición. Esas oraciones violan la Ley de Leibniz (como se ha señalado anteriormente). Por ejemplo, en la oración «Edipo sabía que él mató al hombre que iba en el carro» no podemos sustituir «el hombre que iba en el carro» por «su padre» sin cambiar el valor de verdad de la oración. La solución de Frege es que en contextos gobernados por verbos como «saber» los términos referenciales no tienen ya su referencia habitual, sino más bien su referencia indirecta. Puesto que la referencia indirecta de «su padre» (esto es: su sentido ordinario) y la de «el hombre que iba en el carro» no son la misma, los dos términos no pueden sustituirse uno por otro y, por consiguiente, no se produce violación alguna de la Ley de Leibniz. [39-40]

Frege extendió esta doctrina del sentido y la referencia más allá de los términos singulares: a las oraciones completas. Para identificar el referente de una oración Frege se basaba sobre una idea central de la lógica moderna de acuerdo con la cual se asocia una función con una oración que convierte a un conjunto de palabras en un valor de verdad. Invocando esta idea, Frege consideró el referente de una oración como su valor de verdad. Así pues, todas las oraciones verdaderas se referían a «lo Verdadero», y todas las oraciones falsas, a «lo Falso». Este enfoque de Frege ha tenido un impacto duradero en la semántica formal y figura en argumentos tales como la demostración de Putnam de que la semántica teórico modelística es imposible (ver Lakoff, 1987; Putnam, 1981)[2]. Frege identificó el sentido de una expresión con el pensamiento que expresa. Sin embargo, rechazó cualquier interpretación psicológica de los pensamientos. Mantuvo que la lógica, incluyendo el análisis lógico del lenguaje ordinario, se dirige hacia fenómenos objetivos, no hacia estados psicológicos subjetivos. De este modo los pensamientos eran, para él, entidades objetivas, no estados de una mente individual. Lo que Frege entendía por pensamientos puede comprenderse mejor como aquello a lo que otros filósofos se han referido como

proposiciones, entidades postuladas en tanto que son las que presentan el significado de una oración y son compartidas por oraciones diferentes con el mismo significado (p. ej.; «La nieve es blanca» y «*Schnee ist weiss*»).

Bertrand Russell se encontraba insatisfecho tanto con el tratamiento de esos problemas por parte de Meinong como por parte de Frege y ofreció su teoría de las descripciones como un modo alternativo de tratar con ellos. La teoría de Russell (1905) estaba diseñada para responder no sólo a los problemas descritos por Meinong y Frege, sino también a otros dos adicionales. Uno de ellos ya estaba sugerido en la discusión de la teoría de Frege. Considérese la oración «Jorge IV quiso saber si Scott era el autor de *Waverly*». Russell observa que, aplicando la Ley de Leibniz en este contexto, la oración en cuestión no sólo pierde la importancia cognitiva de la afirmación de identidad (como lo hacía en la oración sobre Venus), sino que también lleva a un enunciado falso. Por ejemplo, si se sustituye «el autor de *Waverly*» por «Scott» se obtiene «Jorge IV quiso saber si Scott era Scott» que presumiblemente es falsa incluso si el enunciado «Jorge IV quiso saber si Scott era el autor de *Waverly*» era verdadero. Un [40-41] principio lógico válido no debería permitirnos inferir un enunciado falso de uno verdadero. El segundo problema adicional de Russell surgía de un principio lógico clásico, el de *tercio excluso*, que afirma que o un enunciado o su contradictorio tienen que ser verdaderos. Pero considérese el enunciado «El actual rey de Francia es calvo». Para evaluar la verdad de este enunciado, miramos la lista de las cosas calvas y, ai no encontrar en ella al actual rey de Francia, concluimos que es falsa. Pero consideremos ahora su contradictoria, «El actual rey de Francia no es calvo». Puesto que el actual rey de Francia no está tampoco en la lista de las entidades no calvas, esta oración es también falsa en una violación aparente de la ley de tercio excluso.

Como alternativa a las explicaciones de Meinong y Frege de esos problemas, Russell avanzó su teoría de las descripciones. De acuerdo con esta teoría, la clase de los nombres se restringe a expresiones que designan directamente individuos que existen de modo efectivo y lo hacen por sí mismos, sin depender del significado de otros términos. (Esta exigencia se impone con la intención de excluir de la clase de los nombres términos como *Sócrates* que, para nosotros, están conectados solamente a su referente por medio de alguna expresión definidora. Sólo tenemos nombres para objetos con los que nos enfrentamos directamente en la experiencia.) Otros términos referenciales, incluyendo muchos nombres aparentes como *Sócrates* y términos descriptivos como la *estrella de la mañana*, se interpretan como descripciones. Así, la expresión «La estrella de la mañana» se analiza como teniendo la forma lógica «el único objeto que tiene la propiedad de ser la última estrella visible por la mañana».

Utilizando este modelo de análisis lógico, Russell propuso disipar todos los problemas mencionados anteriormente. En primer lugar, el enunciado «el cuadrado redondo no existe» se analiza como «No hay ningún objeto que sea a la vez redondo y cuadrado». En este análisis no hay ningún término sujeto que intente referirse al objeto cuya existencia se niega, de modo que el problema se resuelve. En segundo lugar, «Venus es la estrella de la mañana» se analiza como «Existe un único objeto que es la estrella de la mañana y que es Venus». El término *estrella de la mañana* deja de ser un nombre, y la oración se contempla en tanto que atribuyendo la propiedad de ser la estrella de la mañana al objeto nombrado, Venus. Analizada de esta manera, la oración no presenta una afirmación de identidad y, por tanto, no puede convertirse en trivial de la manera que Frege temía. En tercer lugar, «Jorge IV quiso saber si Scott era el autor de *Waverly*» se analiza como «Jorge IV quiso saber si una y sólo una persona escribió

Waverly y si Scott era esa persona». En esta paráfrasis «el autor de *Waverly*» no aparece como un nombre, y de este modo «Scott» no [41-42] puede sustituirlo. Más bien se interpreta que Jorge IV está preguntando si Scott era la persona a la que se aplica el predicado «escribió *Waverly*». Finalmente, «El actual rey de Francia es calvo» se analiza como «Hay una y sólo una persona que es el actual rey de Francia y es calva». Su contradictoria parece ser ahora «No es el caso que [haya uno y sólo un rey de Francia y él es calvo]». (Los corchetes indican que la negación afecta a todo el enunciado y no sólo al primer miembro de la conjunción.) Aunque la primera oración es falsa, su contradictoria es verdadera y no se viola la ley del tercio excluso.

Lo que tanto Meinong como Frege y Russell intentaban hacer en respuesta a estos problemas era articular una teoría del significado. El corazón de todas sus teorías era la noción de referencia: el significado de un término consistía primariamente en el objeto al que se aplicaba. La noción de sentido de Frege y la explicación por parte de Russell de las descripciones se añadían a la explicación referencial para evitar ciertos problemas lógicos con los que la teoría parecía enfrentarse. Ellas constituían, sin embargo, algo adicional y no afectaban al núcleo de la teoría, el concepto de referencia. (Para una discusión adicional de Meinong, Frege y Russell, véase Linsky, 1967.)

La concepción del lenguaje cuyo núcleo consiste en afirmar que funciona refiriéndose a cosas fue desarrollada posteriormente por Ludwig Wittgenstein, particularmente en su *Tractatus Logico-Philosophicus* (1921/1961). La preocupación de Wittgenstein era explicar cómo puede usarse el lenguaje para presentar información sobre el mundo. Los instrumentos de la lógica oracional permitían el análisis de enunciados sobre el mundo en enunciados o proposiciones simples. Esas proposiciones presentaban hechos simples sobre el mundo. El principal interés de Wittgenstein residía en cómo esas proposiciones representaban hechos. Aquí Wittgenstein desarrolló lo que se conoce como la *teoría figurativa del significado*. Su propuesta es que las proposiciones representan rasgos del mundo del mismo modo que lo hacen los dibujos o los mapas. Las líneas y las formas de un dibujo están por las cosas dibujadas y se supone que la relación de las líneas y las formas muestran la relación entre esas cosas. Similarmente Wittgenstein propuso que las palabras de una proposición están por cosas del mundo y las relaciones entre las palabras representan las relaciones entre esas cosas. Cuando el mundo es como la proposición que lo figura, entonces la proposición es verdadera. En esta concepción de cómo el lenguaje describe el mundo, se considera que todos los términos hacen de nombres, de modo que la relación de nombrar es central.

Los análisis referenciales del lenguaje fueron desarrollados adicionalmente por un grupo de filósofos a los que comúnmente se hace [42-43] referencia como los positivistas lógicos. Los positivistas lógicos, que incluyen figuras tales como Carnap, Reichenbach y Hempel, se discuten más detalladamente en Bechtel (en prensa b); aquí sólo menciono un aspecto de sus preocupaciones. Al proponer su teoría de las descripciones, Russell parecía mantener que el lenguaje ordinario podría no manifestar su lógica de la manera más clara posible y podría necesitar una reformulación. Una de las cosas que los positivistas trataban de desarrollar era un lenguaje lógicamente propio que exhibiese claramente la lógica. Con tal lenguaje la gente ya no se despistaría más por rasgos de los lenguajes naturales tales como las expresiones no referenciales. El principal foco de atención de los positivistas lógicos fue el lenguaje de la ciencia. Ellos contemplaban la ciencia como nuestro mayor instrumento para describir verdades e intentaron entender la lógica de la investigación científica y el modo en que el discurso científico

adquiría su significado. Propusieron que el significado de los términos científicos estaba fundado en las experiencias mediante las cuales los científicos podían determinar si los términos eran satisfechos. Tal determinación podría lograrse por observaciones simples o esfuerzos experimentales. La exigencia de que los términos se fundamentasen de esta manera en la experiencia llegó a conocerse como «la teoría del significado como verificabilidad». Una vez que mantuvieron que tal confianza en la verificación era crucial para la ciencia, los positivistas lógicos propusieron extender la exigencia de verificabilidad a otras áreas de la investigación humana y defendieron esa exigencia de modo general para el discurso significativo, Propusieron entonces una modificación del lenguaje ordinario mediante la cual lo expurgaríamos de aquellos términos que carecían de tal verificabilidad. Al desarrollar la teoría de la verificabilidad del significado los positivistas adoptaron el enfoque referencial del lenguaje y lo incorporaron dentro de un análisis de cómo podríamos adquirir conocimiento.

1.3. LA CRÍTICA POSTERIOR DE WITTGENSTEIN A LA TEORÍA REFERENCIAL

En el capítulo anterior vimos que durante algún tiempo Wittgenstein apoyó el análisis referencial del significado. Después de defenderlo en el *Tractatus*, Wittgenstein abandonó la filosofía durante más de una década. Cuando volvió a ocuparse de ella en 1929, comenzó a cuestionarse sus puntos de vista primitivos. Esta revisión culminó en sus *Investigaciones filosóficas*, libro publicado postumamente en 1953. En esta exposición de sus nuevos puntos de vista filosóficos, [43-44] Wittgenstein se centró en la variedad de modos en que es usado el lenguaje y particularmente en el hecho de que puede usarse para hacer más cosas que enunciar hechos. Más bien que buscar el significado de las expresiones lingüísticas en el modo en que las palabras se refieren a objetos, Wittgenstein afirmó que deberíamos centrarnos en cómo usamos el lenguaje. Para capturar la idea de que hay una variedad de usos del lenguaje, Wittgenstein introdujo la idea de que los usos particulares del lenguaje pueden interpretarse como actividades lingüísticas particulares o juegos de lenguaje. Wittgenstein mantuvo que hay una gran variedad de juegos de lenguaje, cada uno con su propio modo de jugarlo y sus propias reglas. En un pasaje de las *Investigaciones* (Wittgenstein, 1953), él nos ofrece la siguiente lista de juegos de lenguaje (que no pretende ser exhaustiva):

Dar órdenes y actuar siguiendo órdenes—
Describir un objeto por su apariencia o por sus medidas—
Fabricar un objeto de acuerdo con una descripción (dibujo)—
Relatar un suceso—
Hacer conjeturas sobre el suceso—
Formar y comprobar una hipótesis—
Presentar los resultados de un experimento mediante tablas y diagramas—
Inventar una historia y leerla—
Actuar en teatro—
Cantar a coro—
Adivinar acertijos—
Hacer un chiste; contarlo—

Resolver un problema de aritmética aplicada—
Traducir de un lenguaje a otro—
Suplicar, agradecer, maldecir, saludar, rezar.
(Wittgenstein, 1953,1, 23.)

Esos diversos juegos de lenguaje las palabras se usan efectivamente de modos diferentes. No se usan siempre para referirse a objetos. De acuerdo con esto, Wittgenstein pensaba que nosotros malentendemos radicalmente el lenguaje ordinario si lo analizamos de manera puramente referencial.

Dolor es uno de los términos que Wittgenstein pensaba que lo malentendíamos si lo tratábamos referencialmente. En el capítulo 5 discuto las propuestas de Wittgenstein respecto de cómo deberíamos entender los fenómenos mentales, pero el hacer alusión a este ejemplo pone de manifiesto uno de los aspectos centrales de su filosofía. Wittgenstein mantiene que muchos errores filosóficos surgen de no prestar atención cuidadosa a la naturaleza de los juegos del lenguaje particulares y a las reglas que los gobiernan. Tales fallos llevan a los filósofos a crear pseudoproblemas. El mismo enunciado de esos problemas [44-45] pone de manifiesto un uso confuso del lenguaje. La tarea propia del filósofo, mantiene él, no es resolver esos problemas, sino disolverlos mostrando cómo se originan a partir de no prestar atención al modo en que el lenguaje se usa realmente. Consideremos el uso de un término como *dolor*. Si no prestamos atención a cómo se usa ese término, podríamos pensar que una oración como «Tengo un dolor» es comparable a la oración «Tengo un gato». Esto podría desorientarnos haciéndonos buscar evidencia de que una persona tiene dolor e intentar caracterizar los dolores como cosas privadas. Pero Wittgenstein nos pide que prestemos atención a las circunstancias en las que usaríamos la expresión «Tengo un dolor». Al usar esta expresión no estamos informando de algo privado, mantiene él, sino que estamos dando expresión a nuestro dolor.

Una de las doctrinas filosóficas sobre el lenguaje que Wittgenstein criticó mantiene que para que un término general (p. ej., *perro* o *libro*) se aplique a un objeto, el objeto debe poseer la esencia apropiada o las propiedades definidoras. La idea de que debe de haber propiedades definidoras para un término general se remonta hasta Sócrates (véase el capítulo anterior) y, desde entonces, ha sido mantenida por muchos filósofos. Wittgenstein niega esta suposición, manteniendo que para muchos términos importantes del lenguaje no podemos especificar propiedades definidoras o esenciales[3]. Esto no es así porque carezcamos de la adecuación necesaria, sino porque el lenguaje no exige que las cosas tengan esencias. Para tratar de convencer a los lectores de esta afirmación, Wittgenstein utiliza el ejemplo del término simple *juego* y defiende que no hay propiedad compartida por todos y sólo los juegos. No hay, por tanto, ninguna propiedad definidora de los juegos compartida por todos y sólo los juegos, sino solamente una gran variedad de similaridades, solapadas unas con otras, entre los distintos juegos:

> Considera, por ejemplo, los procesos que llamamos «juegos». Me refiero a juegos de tablero, juegos de cartas, juegos de pelota, juegos de lucha, etc. ¿Qué hay de común a todos ellos? No digas: «*Tiene que* haber [45-46] algo en común a ellos o no los llamaríamos "juegos"», sino *mira* si hay algo en común a todos ellos. Pues si los miras no verás por cierto algo que sea común a *todos*, sino que verás semejanzas, parentescos y por cierto toda una serie de ellos. Como se ha dicho: ¡no pienses, sino mira!... ¿Son todos ellos «*entretenidos*»? Compara el ajedrez con las tres

en raya. ¿O hay siempre un ganar y perder, o una competición entre los jugadores? Piensa en los solitarios. En los juegos de pelota hay que ganar y perder; pero, cuando un niño lanza la pelota a la pared y la recoge de nuevo, ese rasgo ha desaparecido. Mira qué papel desempeñan la habilidad y la suerte. Y cuan distinta es la habilidad en el ajedrez y la habilidad en el tenis. Piensa ahora en los juegos de corro: Aquí hay el elemento de entrenimiento, ¡pero cuántos de los otros rasgos característicos han desaparecido!...

Y el resultado de este examen reza así: Vemos una complicada red de parecidos que se superponen y entrecruzan. Parecidos a gran escala y de detalle. (Wittgenstein. 1953,1, 66.)

Wittgenstein introdujo la noción de «aire de familia» para describir su punto de vista alternativo acerca de lo que agrupa las cosas en géneros. Al igual que los miembros de una familia humana pueden parecerse entre sí sin que haya una o más características compartidas por todos, Wittgenstein argumentó que las instancias de juegos se parecerían una a otra y, por tanto, formarían una red de vínculos, sin que hubiese una única propiedad compartida por todos los juegos. Éste punto de vista de Wittgenstein ha llegado a ser influyente en la ciencia cognitiva reciente a través de la obra sobre conceptos y categorización de Eleanor Rosch (1975) y otros (véase Smith y Medin, 1981, para un sumario). Rosch rechaza además el punto de vista de que hay condiciones necesarias y suficientes que determinan la pertenencia a una categoría y en lugar de esto explora cómo los miembros de una categoría manifiestan similitud con un ejemplar, (Wierzbicka, 1987, ha desafiado la afirmación de que los términos como juego carecen de un conjunto de propiedades definidoras. Para una discusión adicional de estos problemas, véase Barsalou, en preparación.)

El enfoque del lenguaje del último Wittgenstein es radicalmente diferente del de los filósofos que han afirmado que el lenguaje ordinario tiene que ser reformado debido a sus deficiencias. El enfoque de Wittgenstein representa una versión de aquello a lo que a menudo se hace referencia como *filosofía del lenguaje ordinario*. Este término representa un compromiso con la adecuación del lenguaje ya existente y una necesidad de prestar atención más cuidadosa a cómo este lenguaje se usa de modo efectivo. De hecho Wittgenstein representa una versión radical de la filosofía del lenguaje ordinario en tanto que, cómo él mantiene, los problemas filosóficos surgen «cuando el lenguaje *se va de vacaciones*», esto es, cuando malusamos el lenguaje ordinario, y en tanto que la solución viene no respondiendo los problemas que los filósofos plantean, sino disolviendo los problemas [46-47] filosóficos mediante la apelación a cómo usamos el lenguaje de modo ordinario.

1.4 LA TEORÍA DE LOS ACTOS DE HABLA: AUSTIN, SEARLE Y GRICE

Wittgenstein no fue el único filósofo que miró hacia el lenguaje ordinario como fuente de inspiración. Austin, Searle y Grice han concordado todos ellos con el juicio de Wittgenstein de que, en vez de intentar reformar el lenguaje ordinario, lo que los filósofos deben hacer es prestar atención más cuidadosa a cómo éste funciona. Sin embargo, éstos desarrollaron una perspectiva más bien diferente de la de Wittgenstein en la medida en que subrayaron que el uso del lenguaje es un género de acción y lo

analizaron de acuerdo con ello.

La idea de analizar el lenguaje como un género de acción fue desarrollada por J.L. Austin. En algunas de sus primeras obras, Austin defendía una distinción entre emisiones *realizativas*, tales como dar una orden, y emisiones *constatativas*. Esta última categoría abarcaba las aserciones básicas que habían sido analizadas por Frege, Russell y el primer Wittgenstein. Austin se centró en el primer género de emisiones, que incluían el uso del lenguaje para llevar a cabo acciones tales como ordenar y preguntar y consideró que esos usos constituían acciones. Sin embargo, en la época en que pronunció sus *William James Lectures*, en 1955 (publicadas postumamente como *How to do Things with Words*, 1962a), llegó a tratar todos los actos de habla como acciones —incluyendo las aserciones ordinarias— y, por tanto, como realizativos. Al analizar esos actos, Austin introdujo una distinción entre tres tipos de actos que podrían realizarse al hacer una emisión, actos que él denominó actos *locucionarios*, *ilocucionarios* y *perlocucionarios*. El acto locucionario consiste en hacer enunciados con palabras, donde las palabras se usan con sentidos particulares para hacer referencia a objetos particulares. El acto ilocucionario consiste en la acción que el hablante realiza al hacer la emisión. Esa acción podría consistir en aconsejar o prometer. Para distinguir el acto ilocucionario del significado de las palabras (que es parte del acto locucionario) Austin habla de esos usos diferentes del lenguaje en tanto que incluyendo diferentes fuerzas ilocucionarias. Finalmente, el acto perlocucionario consiste en el efecto que la emisión tiene sobre el oyente. Podría consistir en aburrir al oyente o convencerlo de que lleve a cabo determinada acción[4]. [47-48]

Una vez que distinguimos el acto ilocucionario realizado al decir algo del acto locucionario de emitir las palabras, estamos en una posición en la que nos es posible observar una variedad de maneras en las que un acto puede fallar o ir mal. Por ejemplo, yo puedo emitir las palabras «Prometo darte un martillo» cuando no tengo ningún martillo, siendo entonces irresponsable, desorientador o imprudente. Esto ha llevado a una investigación sobre las condiciones que deben cumplirse para que un acto de habla tenga una fuerza ilocucionaria particular o para que tenga el efecto perlocucionario que se intenta que tenga. Austin comenzó tal investigación, que ha sido continuada más extensamente por John Searie (1969, 1979). Searle, por ejemplo, propuso que, para que una persona pida a otra que haga algo, deben cumplirse las condiciones siguientes: la segunda persona tiene que tener la capacidad de llevar a cabo la acción, y la primera persona tiene que querer que la acción se lleve a cabo, debe creer que su emisión cumplirá este fin y debe de tener razones para que se lleve a cabo. Si alguna de estas condiciones no se satisface, entonces no ha ocurrido acción alguna de pedir.

Los teóricos de los actos de habla se han centrado también en otro rasgo de las acciones realizadas al usar el lenguaje: la cooperación requerida entre los hablantes, Grice (1975) ha articulado cuatro clases de máximas que especifican las maneras en que los hablantes genera] o convencionalmente cooperan en las conversaciones: [48-49]

1. Máxima de Cantidad: Proporciona tanta información como en un contexto se exija, pero no más que la que se exija.

2. Máxima de Cualidad: Proporciona información veraz.

3. Máxima de Relación: Haz que tu contribución sea relevante para el contexto en el que estás hablando.

4. Máxima de Modo: Habla tan claramente como sea posible, evita ambigüedad, di las cosas de la manera más simple posible.

Cuando se violan esas máximas, mantenía Grice, se puede desorientar a la persona al a que se está hablando. Por ejemplo, si sabes que los Reds ganaron el partido pero dices: «Ganaron los Reds o los Pirates» en respuesta a la pregunta «¿Quién ganó el partido?», desorientas a tu auditorio haciéndoles pensar que no sabes quién ganó. Has violado el principio de proporcionar tanta información como se necesita en el contexto. Estos principios no sólo afectan a los efectos perlocucionarios de una emisión, sino también a la fuerza ilocucionaria. Esto se debe al hecho de que, confiando en esas máximas, puedes a menudo querer decir cosas sin decir efectivamente las palabras apropiadas. Por ejemplo, si en respuesta a alguien que dice «Se me ha acabado la gasolina» dices: «Hay una gasolinera a la vuelta de la esquina», puedes estar realizando el acto ilocucionario de decir a alguien donde puede obtener la gasolina que necesita. Pero esto depende de que tu respuesta sea relevante para el contexto y estés dando información máxima. Si sabes que la gasolinera está cerrada, entonces estarías violando la máxima de cantidad y no lograrías realizar el acto ilocucionario de informar a la persona en cuestión de dónde obtener gasolina.

La confianza en esas máximas para realizar actos ilocucionarios genera aquello a lo que Grice se refiere como *implicaturas conversacionales*. Por ejemplo, si al escribir una carta de recomendación para un estudiante un profesor no hace comentario alguno sobre asuntos pertinentes tales como las capacidades escolares del estudiante y se concentra en detalles irrelevantes como la constancia del alumno en la asistencia a las clases, entonces el profesor hace enunciados sobre la actuación del estudiante en cuanto tal y en cuanto que futuro profesional. Sin denigrar explícitamente al estudiante el profesor, sin embargo, lo hace.

Los teóricos de los actos de habla Austin, Searle y Grice, junto con Wittgenstein, intentaron transformar radicalmente la tarea de la filosofía del lenguaje. Más bien que centrarse en el significado de las palabras en el lenguaje, esos filósofos intentaron dirigir el foco de atención de los filósofos hacia la actividad de usar el lenguaje. Durante un gran número de años este enfoque atrajo un considerable interés [49-50] filosófico si bien éste ha ido extinguiéndose a medida que muchos filósofos han vuelto a analizar la estructura formal del lenguaje y a intentar articular la lógica del lenguaje. Las cuestiones sobre cómo se usa el lenguaje han sido absorbidas, sin embargo, por la lingüística como parte de la pragmática (véase Green, en preparación, para más detalles). Además, muchos de los problemas discutidos por esos filósofos, especialmente la distinción de Austin entre fuerzas locucionarias, ilocucionarias y perlocucionarias en los actos de habla, se han convertido en moneda corriente en las investigaciones psicológicas sobre la comprensión del lenguaje y en el trabajo de inteligencia artificial sobre el procesamiento del lenguaje natural.

1.5 ANÁLISIS HOLISTAS DEL SIGNIFICADO: QUINE Y DAVIDSON

Durante el mismo período en que los filósofos del lenguaje ordinario estaban desafiando el enfoque referencial del lenguaje, Quine planteó un género de objeción diferente a ese programa. Quine afirmaba que estaba llevando a cabo el programa de los positivistas lógicos, especialmente el de Carnap, a su conclusión lógica. En un temprano e influyente artículo, Quine (1953/1961a) atacó dos afirmaciones mantenidas por muchos empiristas, que él consideraba que eran desorientadores dogmas que deberían eliminarse del empirismo. Éstos consistían en la suposición de que algunos enunciados eran

analíticamente verdaderos, esto es: verdaderos en virtud de los significados de las palabras (véase el capítulo 1 de este volumen), y que el discurso significativo podía reducirse de una manera sistemática a la experiencia sensorial.

La noción de verdad analítica ha sido particularmente importante para los filósofos de la tradición analítica, cuyo objetivo ha sido descubrir verdades analizando los significados de los términos filosóficamente importantes. Pero Quine argumentaba que no hay definición no circular de analiticidad. Si definimos las verdades analíticas como enunciados verdaderos en virtud de los significados de sus términos, entonces tenemos que definir significado, y Quine argumentaba que esto nos retrotrae a la noción de analiticidad. Quine argumentó que esta incapacidad de definir analítico es un síntoma de un problema más amplio: que las palabras no tienen significados específicos, sino solamente significados en el contexto de toda una red de otras palabras a las que están conectadas en las oraciones que consideramos como verdaderas.

Del mismo modo, Quine argumentaba que los fallos sucesivos en [50-51] el intento de reducir el lenguaje científico a la experiencia sensorial son también síntomas del mismo problema. El remedio que él recomendaba era abandonar la idea de que las palabras o incluso las oraciones tienen significados separados. Más bien, el mantenía que las palabras y las oraciones se entienden mejor en términos de nuestro discurso científico como un todo. Este discurso intenta acomodar nuestra experiencia en el mundo haciendo los ajustes apropiados a lo largo del tiempo. Esto no puede cumplirse si se tienen términos individuales con vínculos fijos con el mundo. En lugar de esto él propuso que debemos contemplar el lenguaje metafóricamente como algo semejante a una fábrica que sólo en su periferia entra en contacto con la experiencia. Quine afirma que esto es una tarea en la que hay gran flexibilidad: podemos introducir modificaciones en diversos lugares en la medida en que hacemos modificaciones adicionales apropiadas en otras partes. A medida que esto sucede, cambian los modos en que las palabras están conectadas entre sí y, por tanto, su significado se altera. (Para una discusión crítica de los puntos de vista de Quine, véase Putnam, 1962,1986. Para una discusión del impacto del ataque de Quine a la analiticidad sobre la filosofía de la ciencia, véase Bechtel, en prensa b.)

Aunque la noción del significado fijo de las palabras se atacaba ya en el desafío de Quine a los dogmas de la analiticidad y de la reducción a la experiencia, Quine (1960) generalizó el ataque cuando desarrolló la tesis de la indeterminación de la traducción. Se centró en la actividad de traducir las emisiones de otra persona en nuestras propias palabras y desarrolló la tesis de que hay siempre una variedad de maneras de hacer esto y que no hay ninguna respuesta determinada a la cuestión de cuál es la traducción apropiada:

> los manuales para traducir un lenguaje a otro pueden establecerse de maneras divergentes, todas ellas compatibles con la totalidad de las disposiciones del habla, aunque incompatibles entre sí. En incontables lugares divergirán al dar, como sus respectivas traducciones de una oración de un lenguaje, oraciones del otro lenguaje que no están entre sí en ningún género plausible de equivalencia, ni siquiera laxo. (1960, p. 27.)

Quine comienza su defensa de esta tesis considerando un caso de traducción radical donde nos enfrentamos a un lenguaje tan remoto del nuestro que no se han desarrollado manuales de traducción

estándar. A continuación intenta convencernos de que la misma moraleja puede ser extraída cuando tratamos con otros hablantes de nuestro lenguaje o incluso con nuestra propia habla pasada. Considera que la tarea de entender o interpretar las palabras que alguien emite es meramente una operación consistente en traducirlas a nuestras [51-52] propias palabras. Tanto en el caso foráneo como en el doméstico, su afirmación es que no hay fundamentos científicamente aceptables (esto es: basados en evidencia sensorial o empírica) para insistir en que una traducción es más correcta que otra. Es ésta una tesis radical. Quine no está haciendo observar meramente que hay una carencia de correlación perfecta entre los lenguajes de modo que no podemos identificar siempre el modo correcto de correlacionar expresiones entre ellos. Más bien está diciendo que siempre habrá interpretaciones alternativas, radicalmente diferentes, de lo que se dice incluso cuando el lenguaje es el de uno mismo. Así podemos considerar que un hablante (incluyendo uno mismo) está diciendo cosas diferentes e inconsistentes dependiendo de qué traducción adoptemos y no hay respuesta alguna a la pregunta de cuál es la correcta (Quine, 1970).

La argumentación de Quine a favor de la tesis de la indeterminación descansa sobre otras dos tesis que él llama la *subdeterminación de las teorías* y la *inescrutabilidad de la referencia*. Presento solamente el argumento a favor de la subdeterminación de las teorías. La tesis de la subdeterminación de las teorías mantiene que en ciencia se pueden construir siempre teorías alternativas que estén de acuerdo con los mismos datos empíricos y que incluso no sea posible decidir, cuando todos los datos posibles se han recogido, entre esas teorías (Quine, 1960, 1970). Quine defiende esta tesis sobre bases empiristas. Sólo permite que la evidencia sensorial decida disputas teóricas, pero hace observar que las teorías científicas hacen afirmaciones que van más allá de la evidencia. La tesis de la subdeterminación mantiene simplemente que dos teorías pueden diferir solamente en las áreas que van más allá de la evidencia, de modo que la evidencia no puede establecer cuál de ellas es la correcta. Quine convierte la tesis de la subdeterminación en un apoyo de la tesis de la indeterminación haciendo que nos imaginemos a nosotros mismos intentando traducir la teoría de algún otro en un dominio para el que poseemos dos teorías subdeterminadas. Quine afirma simplemente que podríamos traducir la teoría de esa persona en cualquiera de nuestras teorías subdeterminadas y no habría nada que contase en favor de una traducción sobre la otra. Así pues, tenemos dos traducciones y no tenemos evidencia alguna sobre cuya base podamos decidir cuál es la correcta (Quine, 1970).

Si las palabras de una persona en un lenguaje tuvieran significados específicos, tal determinación no surgiría. La conclusión que extrae Quine es, sin embargo, que no tenemos evidencia para tales significados y que, por tanto, debemos abandonar la idea de que las palabras tienen significados específicos. Además, afirma él, no hay [52-53] significados o proposiciones en las cabezas de los usuarios del lenguaje que determinen cómo debemos interpretar su lenguaje[5]. Como resultado de esto Quine no contempla el uso del lenguaje como una actividad mental peculiar. Es más bien un fenómeno de la naturaleza que deberían explicar los científicos. (Quine, 1973, propone un análisis protocientífico.) Este propósito consistiría simplemente en articular la estructura lógica del lenguaje y en mostrar cómo se relaciona con el mundo en que existe el hablante (véase capítulo 3). Sobre la base de su propio análisis lógico, Quine argumenta que algunas formas del discurso humano no están estructuradas adecuadamente para su uso en la investigación científica. Por ejemplo, él ha argumentado que el discurso modal del tipo discutido en la sección que sigue, lo mismo que la cita indirecta (donde intentamos capturar lo que

alguien dijo en palabras diferentes), son modos de discurso pobremente trabajados que deberíamos abandonar al menos cuando estamos haciendo ciencia y queremos desarrollar explicaciones verdaderas de la naturaleza (Quine, 1960).

Aunque Quine propone una filosofía del lenguaje que no deja lugar para una explicación del significado, Donald Davidson, filósofo que ha sufrido una influencia significativa por parte del argumento de la indeterminación de Quine, ha intentado, a pesar de todo, articular una teoría del significado dentro de la perspectiva quineana básica. Para hacerlo, Davidson invoca el análisis de Tarski de la verdad para los lenguajes formales. Tarski (1944/1952, 1967) propuso enunciar la condición de verdad para una oración dada en un lenguaje formal en términos de V-oraciones tales como la siguiente:

«La nieve es blanca» es verdadera si y sólo si la nieve es blanca.

Para ver que este enunciado es algo más que trivial, debe reconocerse una diferencia fundamental entre las dos ocurrencias de las palabras [53-54] «la nieve es blanca». En la primera ocurrencia las comillas nos dicen que estamos nombrando y, por ende, refiriéndonos a la oración «la nieve es blanca», mientras que en la segunda ocurrencia estamos usando la oración para designar el hecho que haría verdadera esta oración. Técnicamente decimos que la segunda ocurrencia está enunciada en un lenguaje diferente, el lenguaje objeto, del que está enunciada la primera ocurrencia, que está en el metalenguaje (un lenguaje usado para hablar sobre el primer lenguaje). Las V-oraciones de Tarski pueden interpretarse como presentando una versión de la teoría de la verdad como «correspondencia» de acuerdo con la cual una oración es verdadera si se corresponde con cómo las cosas son efectivamente. Una definición adecuada de la verdad tiene que tener como consecuencias lógicas, para Tarski, todas las V-oraciones para el lenguaje. Tarski demostró que para los lenguajes formales que reúnen determinadas condiciones es posible producir una tal definición de verdad, pero que esto no es posible para lenguajes ordinarios como el castellano.

Aunque la meta de Tarski era definir verdad, Davidson (1967) considera la verdad como un primitivo y usa el esquema de Tarski para dar cuenta del significado. Así, para Davidson, identificamos el significado de una oración estipulando qué sería el caso si la oración fuese verdadera. Si estamos intentando enunciar el significado de una oración en nuestro propio lenguaje, entonces, como sucede en la V-oración que se acaba de dar, usaremos nuestro lenguaje como el metalenguaje en el que enunciar las condiciones de verdad. Pero, cuando estamos dando el significado para oraciones de un lenguaje foráneo, nombraremos la oración del lenguaje foráneo y enunciaremos las condiciones de verdad en castellano, como sucede en el ejemplo siguiente:

«*Schenee is weiss*» es verdadera en alemán si y sólo si la nieve es blanca.

La tarea de una *teoría del significado*, en tanto que una definición de verdad, es generar V-oraciones para todas las oraciones del lenguaje. Para hacer esto no podemos simplemente enunciar las V-oraciones para cada oración del lenguaje, puesto que habrá un número infinito de tales oraciones. Más bien necesitamos desarrollar un procedimiento recursivo que muestre cómo construir una V-oración para cualquier oración dada.

Al desarrollar esta explicación del significado que descansa solamente sobre condiciones de verdad, Davidson afirma que permanece dentro de las constricciones quineanas. Además, él apoya también el [54-55] holismo de Quine sobre el significado. En la práctica nos enfrentamos a la tarea de adscribir significado cuando necesitamos interpretar o traducir lo que alguien está diciendo. Lo que estamos haciendo es intentar figurarnos lo que sería verdadero si lo que se está diciendo lo fuese. Al igual que Quine, Davidson mantiene que en estas circunstancias no tenemos un criterio independiente mediante el que fijar el significado de las palabras. Para poder iniciar el procedimiento, mantiene Davidson, tenemos que suponer que la persona en cuestión está diciendo, al menos durante la mayor parte del tiempo, lo que nosotros consideraríamos también que es la verdad y que también está diciendo lo que nosotros diríamos. Davidson caracteriza esto como un *principio de caridad*: interpretamos a la otra persona como diciendo tantas cosas verdaderas como sea posible (Davidson, 1973, 1974a, 1975). Al adoptar este principio intentamos construir una teoría de la interpretación que aparee las oraciones de otra persona con las oraciones nuestras que sean equivalentes en valores de verdad. Solamente si encontramos puntos en los que nuestra teoría generativa haga encajar mejor oraciones que consideramos falsas con oraciones que la otra persona considera que son verdaderas, reconocemos que la otra persona puede creer falsedades. Una motivación para aceptar este principio es que interpretamos las palabras de otro para adquirir información. Podemos obtener información sólo si desarrollamos un esquema que interpreta a la otra persona de tal manera que esté diciendo la verdad la mayor parte del tiempo. Davidson extrae moralejas muy fuertes de este principio de caridad. Por ejemplo, niega que podamos entender la idea de que otra persona tenga esquemas conceptuales o modos de entender el mundo que sean radicalmente diferentes de los que nosotros usamos[6]. Su razón es que no trataríamos los enunciados de la persona en cuestión como si constituyeran un esquema conceptual a menos de que pudiésemos interpretarlos, y por el principio de caridad tenemos que interpretar la mayor parte de esos enunciados como verdaderos (Davidson 1974b). [55-56]

1.6 DISCURSO MODAL, SEMÁNTICA DE MUNDOS POSIBLES Y TEORÍAS CAUSALES DE LA REFERENCIA: KRIPKE Y PUTNAM

Los análisis filosóficos del lenguaje que he examinado hasta ahora se han concentrado en él lenguaje extensional, donde los términos pueden tratarse como haciendo referencia a objetos efectivamente existentes y las oraciones pueden contemplarse como adscribiendo propiedades o relaciones a esos objetos. En tales contextos, la Ley de Leibniz sanciona la sustitución de un término por otro con el mismo referente sin cambiar la verdad de una oración. Quine y Davidson son dos filósofos contemporáneos que han argumentado de manera vociferante a favor de limitar el discurso significativo a contextos extensionales y a favor de rechazar contextos no extensionales como lingüísticamente sospechosos. Pero los contextos no extensionales son comunes en el habla corriente. Uno de tales contextos incluye el uso de verbos como «saber» y «creer», que ya hemos encontrado al discutir a Frege y que discuto en capítulos posteriores. Otra clase común de oraciones no extensionales son aquellas que contienen las que comúnmente se llaman palabras *modales*, tales como «necesariamente», «tiene que», «posiblemente» o «puede». Considérese la oración:

Era posible que Nixon podría no haber sido Presidente.

Si sustituimos Nixon por el término correferencial *el Presidente número 37*, obtenemos:

Era posible que el Presidente número 37 podría no haber sido Presidente.

Aunque la primera oración parece ser verdadera, la segunda no[7]. [56-57]

En castellano ordinario hay una gran variedad de inferencias aparentemente válidas que usan términos modales. Éstos, sin embargo, no resultan sancionados por los principios del cálculo de predicados ordinario, que está ligado al lenguaje extensional. Durante este siglo se han avanzado un buen número de propuestas para modificar el conjunto de axiomas que gobiernan el cálculo de predicados para acomodar esas inferencias (Carnap, 1956; Church, 1943). Estas propuestas no estaban, sin embargo, acompañadas por teorías semánticas apropiadas para explicar los operadores modales. Esta deficiencia se remedió cuando Kripke (1963) desarrolló una interpretación teórico-modelista de varios conjuntos de axiomas para la lógica modal. A continuación, Kripke (1971-1972), Donnellan (1972), Putnam (1973,1975b) y otros han intentado mostrar cómo el análisis formal de los enunciados modales puede servir para arrojar luz sobre problemas básicos de la filosofía del lenguaje tales como el del significado de los nombres propios y comunes.

El problema para entender un enunciado modal como «Reagan podría no haber sido elegido Presidente» de una manera extensional es que es un enunciado contrafáctico. Nos pide que contemplemos cómo las cosas podrían haber sido diferentes. Claramente, no podemos juzgar la verdad y figurarnos el significado de tales enunciados determinando si Reagan fue elegido presidente. Un modo común de representar lo que tales enunciados están afirmando es invocar la idea de *mundo posible*. Esta idea se retrotrae en última instancia a Leibniz, que representaba a Dios como contemplando combinaciones lógicas diferentes de individuos y eligiendo este mundo como el conjunto más vastamente compatible (invitando así a Voltaire a hacer el comentario satírico de que este es el mejor de los mundos posibles). La noción de mundo posible se usa para explicar la lógica modal invitándonos a pensar en universos alternativos que se definen en términos de cambios específicos a partir de este universo. Consideramos entonces cómo las otras cosas serían diferentes bajo esas situaciones. Así podríamos considerar el mundo en el que Adolf Hitler hubiese sido un aborto en lugar de haber nacido y a continuación rellenar el resto del escenario para ese mundo. Si invocamos esta ficción de los mundos posibles, entonces estamos en posición de decir lo que hace que un enunciado modal sea verdadero o falso. Una afirmación de que un objeto necesariamente tiene una propiedad es verdadera justamente en el caso en que tiene esa propiedad en todo mundo posible en que el objeto existe. Así, Ronald Reagan era necesariamente un actor es verdadera si, en todo mundo posible en el que Ronald Reagan existe, él es también un actor. Puesto que hay un mundo posible en el que existe y no es un actor, el enunciado es falso. [57-58]

Al interpretar las afirmaciones modales en términos de mundos posibles, Kripke avanzó su argumento de que los nombres son lo que él llamó *designadores rígidos*, y no equivalentes a ninguna descripción que podríamos usar para seleccionar el referente. El argumento descansa sobre nuestra aceptación de su intuición de que podemos contemplar la posibilidad de que la persona o cosa en cuestión no tuviese las propiedades que supuestamente usamos para identificarla. Por ejemplo: podemos seleccionar a Richard

Nixon como la persona que fue el Presidente número 37 de los Estados Unidos, pero a continuación podemos contemplar la posibilidad de que no hubiese sido elegido nunca presidente. Así, Kripke afirma, el nombre no es idéntico a la descripción. (Véase Linsky, 1977, para un rechazo de este argumento.) El nombre selecciona la persona u objeto mismo, sin tener en cuenta qué propiedades podría haber tenido esa persona en el mundo que se está considerando. No es necesario que exista la persona Richard Nixon, pero, en cualquier mundo en el que Nixon exista, el designador rígido «Richard Nixon» selecciona a esa persona.

La tesis de Kripke equivale a la afirmación de que los nombres propios no tienen sentidos fregeanos, sino sólo un referente. Éste fue un punto de vista mantenido incluso con anterioridad a Frege por J. S. Mili (1846) y en sí mismo puede no parecer terriblemente sorprendente. Pero Kripke y otros, que avanzan este enfoque para comprender los contextos modales, avanzan también una tesis similar sobre los nombres comunes que se refieren a «géneros naturales» como carbón u oro. Esos términos funcionan igualmente como designadores rígidos, seleccionando objetos particulares sin tener en cuenta las propiedades que usamos para identificarlos, y de este modo carecen también de sentidos y poseen sólo referentes. El argumento a favor de esta tesis es muy similar al de los nombres propios. Puesto que es posible que el objeto en cuestión podría no tener la propiedad que asociamos con el nombre (p. ej., el oro podría no ser amarillo en algún mundo posible), la propiedad no puede determinar la referencia. Puede ser sólo una muletilla usada en este mundo para transmitir la referencia a alguna otra persona, pero, una vez que se fija la referencia, la propiedad ya no figura más como parte del significado del nombre.

Una vez que se ha rechazado el punto de vista de que ni los nombres propios ni los nombres comunes están asociados con propiedades que sirvan para seleccionar sus referentes, Kripke y otros defensores de los enfoques modales han llegado a ofrecer una concepción diferente de cómo esos nombres están ligados con sus referentes. Han avanzado lo que se llama una teoría causal de los nombres. La idea es que los nombres están ligados a sus referentes por una cadena [58-59] causal. Por ejemplo, en una ceremonia de bautismo podría asignarse un nombre a una persona. Todo uso ulterior de ese nombre para esa persona se retrotrae al nombrar original. Similarmente, cuando alguien encuentra por vez primera una instancia de un género natural como un trozo de oro, esa persona podría asignar el nombre «oro» a ese género. El uso ulterior del nombre para substancias de ese género estará ligado a él mediante esa cadena causal. (No es relevante para el *significado* del término, de acuerdo con los teóricos de esta tradición, que para identificar instancias subsiguientes del género natural necesitemos basarnos en procedimientos de identificación.)

La teoría causal se contempla por sus proponentes como un desafío directo a una gran variedad de puntos de vista tradicionales sobre el significado. En particular, es un desafío a la idea fregeana de que los términos tienen tanto sentido como referencia. Se contempla también como un desafío a la alternativa wittgensteiniana de la idea fregeana. Wittgenstein (como se ha discutido previamente) propuso que, aunque pueda no haber rasgos definitorios compartidos por todos los objetos a los que hace referencia un término, puede haber un aire de familia entre ellos. Los teóricos causales niegan que haya ningún conjunto tal de propiedades que determine el significado de tales términos. Más bien el término se aplica directamente al objeto, puesto que la conexión se estableció por el acto inicial de nombrar al objeto[8].

Esos intentos de explicar lo que se quiere decir mediante las oraciones modales invita a plantear una

pregunta sobre cómo reconocemos objetos en mundos posibles. ¿Cómo determinamos qué entidad es en otro mundo posible Richard Nixon o un trozo de oro? Kaplan (1967) llamó a esto el problema de la *identidad transmundana*[9]. Kripke (1972) responde que el mero planteamiento de esta pregunta representa un error fundamental. Los mundos posibles no son cosas que identifiquemos en primer lugar y después determinemos cómo se corresponden sus habitantes con los del mundo actual. Los mundos [59-60] posibles se *estipulan*, no se descubren. Estipulamos qué individuos existen en el mundo posible y qué propiedades tienen. Por consiguiente, jamás necesitamos plantear cuestiones acerca de qué individuo corresponde a un individuo de nuestro mundo. Comenzando con Richard Nixon, decidimos si existe o no en el mundo posible que estamos contemplando y si, existe, entonces le atribuimos todas sus propiedades *esenciales* (que son típicamente diferentes de aquellas que usamos para identificar a Richard Nixon) y cualesquiera otras propiedades que consideramos que tiene en el mundo posible.

Aunque este enfoque evita el problema de especificar las relaciones de identidad transmundana, provoca otra objeción concerniente a las propiedades esenciales que tienen que atribuirse a cualquier individuo en un mundo en el que el individuo existe. El punto de vista de que algunas de las propiedades de una entidad le son esenciales de modo que, si careciese de esas propiedades, no sería la misma entidad se conoce como *esencialismo*. Para identificar esas propiedades, Kripke, Donnellan y Putnam se basan de manera importante en sus intuiciones sobre lo que hace que un objeto sea el objeto que es. En el caso de los seres humanos, Kripke considera que su origen constituye su propiedad esencial. Así, aunque Nixon hubiera sido un luchador de *sumo*, no podría haber nacido de padres diferentes. En el caso de los elementos químicos, como el oro o el agua, Putnam (1975b) mantiene que lo esencial es su composición molecular. Así el agua es H_2O en cualquier mundo en el que exista, aunque podría diferir de nuestra agua por lo que respecta a otras propiedades. En el caso de los artefactos, Kripke considera la materia de la que están hechos como crucial para su identidad. Así, Kripke argumenta que un atril que está hecho de manera efectiva de un cierto trozo de madera no podría estar hecho de agua congelada del río Támesis. Algunas intuiciones no son compartidas, sin embargo, por todo el mundo. Por ejemplo, alguien podría afirmar que lo que parece crucial para que alguien sea Richard Nixon es su apariencia física o que sea un político. Si careciese de esas propiedades una persona no podría ser simplemente Richard Nixon.

Es difícil ver cómo los argumentos pueden establecer lo que es esencial para que algo sea la entidad o el género de entidad que es de hecho. El hecho de que los juicios sobre lo que es esencial parezcan descansar sobre nada más que las intuiciones de algunos hablantes es una razón por la que algunos filósofos han considerado como problemática toda la tarea de evaluar las afirmaciones modales. El argumento para interpretar los nombres como designadores rígidos, sin propiedades o sentido alguno pegados a ellos, depende entonces fuertemente de esos argumentos modales. Así, si se rechazan los [60-61] contextos modales y el esencialismo, como hacen Quine y Davidson, entonces uno puede estar perfectamente de acuerdo en asociar nombres con descripciones o incluso en eliminar completamente los nombres, como Quine (1960) propone. Por otra parte, el aceptar el discurso modal y el diseñar una semántica para él, parece exigir una concepción radicalmente diferente de los nombres y una clara distinción entre nombres y descripciones. (Para discusiones adicionales de esos problemas, ver Lewis, 1983b; Linsky, 1977.)

1.7 RESUMEN

Hemos estado haciendo un repaso de diferentes análisis filosóficos del lenguaje. El análisis referencial adoptado por Frege y Russell ha sido criticado de diversas maneras por Wittgenstein y los teóricos de los actos de habla, que argumentan que para entender un lenguaje debemos mirar cómo se usa. Pero otros filósofos contemporáneos han adoptado versiones modificadas de los análisis referenciales de Frege y Russell. El carácter extensional de las teorías referenciales ha sido mantenido por Quine y Davidson, que han desafiado otros rasgos tales como la introducción de los sentidos por parte de Frege y la teoría de las descripciones de Russell. Tanto Quine como Davidson rechazan la idea de un significado para las palabras, además de la referencia, y colocan la asignación de la referencia dentro de una perspectiva holista. Del mismo modo, Kripke y Putnam atacan la noción fregeana de sentido, pero también rechazan el extensionalismo del viejo enfoque referencial: Han propuesto una teoría causal mediante la que los nombres están ligados causalmente a sus referentes y mantienen este vínculo a través de los mundos posibles.

En los capítulos que siguen será obvio que esas teorías del lenguaje tienen implicaciones para las teorías sobre la mente. (Para dos discusiones contemporáneas de problemas de filosofía del lenguaje que hacen explícitas sus implicaciones para teorías de la mente, véase Lycan, 1984, y Pollock, 1982.) Debido a esas conexiones entre teorías del lenguaje y teorías de la mente, en los capítulos que siguen hay numerosas referencias retrospectivas al material introducido aquí. Pero ahora es tiempo ya de entrar en la discusión de un problema central de la filosofía de la mente, el problema de si los fenómenos mentales se distinguen de los fenómenos físicos como resultado de poseer una propiedad conocida como *intencionalidad*. [61-62]

3. EL PROBLEMA DE LA INTENCIONALIDAD

3.1 INTRODUCCIÓN

Un estado mental típico, por ejemplo una creencia, es generalmente *sobre* algo. Puedes creer, por ejemplo, que Hawai es un lugar hermoso, en cuyo caso tu creencia es *sobre* Hawai. Esta característica de ser sobre algo es lo que los filósofos llaman *intencionalidad* [1]. Sin embargo, muchos filósofos consideran la intencionalidad como un rasgo que diferencia los estados mentales de otros fenómenos de la naturaleza. La aspiración de este capítulo es introducir el fenómeno de la intencionalidad y discutir por qué algunos filósofos lo han contemplado como si presentase un obstáculo para desarrollar explicaciones científicas de los fenómenos mentales. En el capítulo 4 volveremos sobre algunas estrategias que otros filósofos han propuesto para explicar la intencionalidad de una manera científicamente aceptable.

Antes de examinar algunos de los criterios más explícitos que se han ofrecido para identificar la intencionalidad de los estados mentales, podemos aislar la idea básica de que la intencionalidad se refiere a la capacidad de los estados o eventos mentales de ser sobre otros objetos o eventos. En la creencia sobre Hawai que se acaba de mencionar, Hawai y la belleza putativa de Hawai son objetos de tu creencia. Siguen siendo los objetos de tu creencia incluso si nunca has estado en Hawai y ahora estás muy lejos de esa tierra. Es relativamente fácil ver, al menos en términos generales, cómo se diferencia este rasgo de los estados mentales de muchos otros estados o eventos de la naturaleza [2]. Los estados ordinarios del mundo, tales como una [62-63] lámpara que está sobre una mesa, no son sobre nada. La lámpara puede ser afectada causalmente por otros objetos, y puede causar cambios en otras entidades de la naturaleza, pero no tiene estados que sean sobre otras cosas en ningún sentido similar al modo en que la gente tiene creencias sobre diversos objetos. La capacidad de ser sobre otros estados es verdadera no solamente de las creencias, sino de toda una hueste de otras actividades mentales, tales como desear, temer, dudar, esperar, planear. Si esperas obtener la cátedra, entonces tu esperanza es *sobre* la cátedra. Mediante tus actividades mentales te conectas con otros estados de la naturaleza, pero no en ningún sentido lisa y llanamente causal. Así puedes tener una creencia sobre un estado de cosas (p. ej., obtener una cátedra) que no está producida causalmente por ese estado de cosas, y puedes tener un deseo de un estado de cosas sin que ese deseo te conduzca a llevar a cabo ninguna acción para producirlo.

Partiendo de esta caracterización informal de la intencionalidad, varios filósofos han intentado desarrollar caracterizaciones más formales que sirvan también para mostrar lo que distingue a los fenómenos intencionales de los que son no intencionales. Dos de las más prominentes se deben al filósofo del siglo XIX Franz Brentano y, más recientemente, a Roderick Chisholm.

3.2 LA EXPLICACIÓN DE BRENTANO DE LA INEXISTENCIA INTENCIONAL

Brentano (1874/1973) enfocó su atención en el hecho de que las cosas o eventos a las que se hace referencia en los estados mentales no necesitan ser reales. Podemos tener creencias u otros estados

mentales sobre objetos no existentes. Por ejemplo, alguien podría creer que los unicornios tienen sólo un cuerno, o un niño podría esperar que Papá Noel le va a traer regalos maravillosos. Aunque ni los unicornios ni Papá Noel existan, Brentano afirma que, con todo, se presentan a la persona en cuestión en los estados mentales. Brentano arguye que todo estado mental —no solamente aquellos que normalmente se considera que incluyen presentaciones (oír un sonido, ver un objeto coloreado, sentir calor o frío), sino también los juicios, recuerdos, inferencias, opiniones, etc.— incluye un objeto u objetos [63-64] que se presentan o aparecen al sujeto. Afirma también que el hecho de que los estados mentales incluyan tales presentaciones de cosas constituye su intencionalidad y es lo que los distingue de todos los fenómenos físicos:

> Todo fenómeno mental se caracteriza por lo que los escolásticos de la Edad Media llamaron la in-existencia intencional (o mental) de un objeto, y que podríamos llamar, aunque no de manera completamente ambigua, la referencia a un contenido, la dirección hacia un objeto (que no ha de entenderse aquí como significando una cosa) o hacia una objetividad inmanente. Todo fenómeno mental incluye algo como objeto dentro de sí mismo, aunque no todos lo incluyen de la misma manera. En la presentación se presenta algo, en el juicio se afirma o se niega algo, en el amor es amado, en el odio odiado, en el deseo deseado y así sucesivamente. Esta in-existencia intencional es característica exclusivamente de los fenómenos mentales. Ningún fenómeno físico exhibe algo parecido a esto. Podemos, por tanto, definir los fenómenos mentales diciendo que son aquellos fenómenos que contienen un objeto intencional dentro de ellos mismos. (Brentano, 1874/1973, p. 88.)

Para Brentano la intencionalidad de los estados mentales no sólo los distingue de los estados puramente físicos; arruina también cualquier intento de estudiar los estados mentales usando las herramientas de la ciencia física. Así pues, el tratamiento de Brentano de la intencionalidad proporciona un apoyo para el punto de vista dualista de que la mente es distinta del cuerpo (para más cuestiones sobre el dualismo, ver el capítulo 5).

El pasaje citado de Brentano ha sido el foco de muchas controversias. De acuerdo con una interpretación, adoptada por el discípulo de Brentano Meinong, Brentano se compromete con una clase de «objetos» (esto es: objetos de pensamiento) que existen incluso cuando no hay objetos en el mundo físico que les correspondan. Así pues, cuando pienso en algo, por ejemplo en el helado perfecto, tiene que haber un objeto particular hacia el que se dirige mi pensamiento. Pero, puesto que no hay ningún objeto real que encaje con esta descripción, el objeto de mi pensamiento tiene que ser un género peculiar de objeto mental [3]. Para acomodar esta interpretación, Meinong (1904/1960) introdujo una distinción entre el *Sosein* (el ser o subsistencia) de un objeto y su *Sein* (existencia). Los objetos que no existen efectivamente, como las montañas de oro o los cuadrados [64-65] redondos, tienen con todo una subsistencia. Es la subsistencia del objeto lo que constituye el objeto intencional del pensamiento. Así, sí digo del cuadrado redondo que es redondo, estoy hablando realmente del cuadrado redondo subsistente y no sobre ningún objeto real.

Aunque esto proporciona una interpretación plausible del pasaje citado de Brentano, Brentano mismo la desautorizó de hecho puesto que se dio cuenta de que llevaba a serios problemas. La dificultad fue

expuesta claramente por Frege. En el capítulo anterior observamos que Frege (1982) introdujo la distinción entre el sentido de una expresión y su referente. El sentido representaba el modo de presentación del objeto (p. ej., caracterizaría los rasgos del objeto). Aunque podría pensarse que los sentidos de Frege pueden servir como objetos intencionales [4], Frege reconoció que, si consideráramos al discutir los objetos de pensamiento que los sentidos son objetos de pensamiento, entonces estaríamos comprometidos a hacer lo mismo cuando discutimos los objetos efectivos. La razón es que no hay nada en el estado mental mismo que distinga los casos en los que estamos pensando sobre objetos efectivos de aquellos en los que estamos pensando sobre objetos no existentes. Esto nos lleva a la indeseada consecuencia de que todo nuestro discurso es sobre sentidos u objetos intencionales y no sobre los objetos del mundo. (Para una discusión adicional de este problema y del tratamiento que Brentano hace de él, véase Chisholm, 1967; Follesdal, 1982; Husserl. 1913/1970, 1929/ 1960,1950/1972.)

Al señalar el hecho de que los estados mentales pueden dirigirse hacia objetos o eventos no existentes, Brentano estableció una difícil tarea para los pensadores siguientes. El hecho de que los estados mentales se dirijan a objetos, y el hecho de que esos objetos no existan siempre, hace difícil dar cuenta de la intencionalidad de los estados mentales. Parece que estamos comprometidos con las afirmaciones inconsistentes de que, por una parte, los estados intencionales incluyen una relación con un objeto y de que, por otra parte, el objeto con el que podríamos esperar poner en relación los estados intencionales no necesita existir. Una relación requiere dos objetos, y con todo, por lo que respecta a los estados intencionales, puede no haber un segundo objeto [5]. Sin embargo, es ésta una tensión para la que no [65-66] hay una resolución fácil, puesto que, como Richardson (1981) ha argumentado, no podemos realmente sacrificar ninguna de las dos afirmaciones si hemos de tratar adecuadamente con los fenómenos intencionales:

> Por una parte, si mantenemos que hay objetos no existentes aunque reales que son los objetos de nuestro pensamiento, estamos obligados a admitir que ninguno de nuestros objetos [los objetos sobre los que tenemos creencias, etc.,] está en el mundo real. Se nos impide pensar en lo concreto. Por otra parte, si admitimos que los actos mentales no son realmente relaciónales, somos llevados a la conclusión de que los actos mentales no pueden realmente dirigirnos hacia objetos del mundo (ni tampoco de fuera de él). Nuestro pensamiento no nos relaciona con el mundo. En cualquier caso, tales actos difícilmente pueden contemplarse como intencionales (pp. 177-178).

Desde la perspectiva de la ciencia cognitiva moderna, podría suponerse que el problema de la intencionalidad se podría resolver postulando representaciones como los objetos de los estados mentales y, por consiguiente, como los objetos de pensamiento. Aunque las representaciones, como se discute más tarde, pueden desempeñar un papel importante al explicar cómo es posible la intencionalidad, no pueden desempeñar el papel para el que Brentano parece estar postulando objetos intencionales. La razón puede apreciarse si nos concentramos en las creencias verídicas. En tales casos queremos decir que nuestras creencias son *sobre* el objeto o estado de cosas efectivamente existente en el mundo. Mas, si hacemos de las representaciones los objetos de nuestras creencias en el caso de falsas creencias, un razonamiento similar exige que consideremos las representaciones como los objetos de creencia en el caso de las

creencias verídicas. Pero esto no logra capturar el importante elemento para el que se introdujo en primer lugar el término *intencionalidad*, a saber: la idea de que el objeto de nuestros estados mentales son a menudo cosas [66-67] externas a nosotros. Si adoptamos la herramienta de las representaciones mentales, tenemos aún que explicar cómo algunos de nuestros estados mentales tienen éxito al conectar con cosas del mundo mientras que otros no logran hacer esto. Es esta conexión, que puede o no puede ocurrir, la que hace que esas representaciones sean de algo y, por tanto, intencionales.

3.3 EL CRITERIO LINGÜÍSTICO DE CHISHOLM DE LA INTENCIONALIDAD

El criterio de Brentano para la intencionalidad parece conducir a un matorral metafísico al plantear cuestiones sobre el *status* de los estados intencionales. Muchos filósofos de habla inglesa, particularmente en la primera mitad de este siglo, han buscado evitar tales cuestiones metafísicas centrándose no sobre los fenómenos del mundo, sino sobre el lenguaje en el que se hacen las afirmaciones sobre el mundo. En particular han intentado mostrar cómo podríamos clarificar y resolver muchos problemas científicos y filosóficos presentando nuestras afirmaciones en términos de la lógica simbólica. Sin embargo, nuestro lenguaje para describir los estados mentales parece introducir algunas peculiaridades lógicas que llevan a Roderick Chisholm, entre otros, a proponer que podríamos identificar los estados intencionales en términos de las peculiaridades lógicas de las oraciones que se refieren a ellos.

Hay dos aspectos importantes de la lógica simbólica moderna que, para los propósitos de la presente discusión, necesitamos tener presentes. El primero de ellos es que la lógica es *veritativo-funcional*. Esto significa que la verdad de cualquier oración que está compuesta de otras oraciones (p. ej., «Hoy es martes o está nevando») puede averiguarse simplemente conociendo el valor de verdad de las oraciones componentes. Un segundo rasgo importante es que la lógica simbólica es *extensional*. Como vimos en el capítulo anterior, esto significa que la verdad de una expresión depende solamente de aquello a lo que la expresión se *refiere* (su *extensión*), no de su significado (*intensión*). Como hemos hecho notar, el discurso extensional obedece a la Ley de Leibniz, que permite la sustitución en un enunciado de un término por otro término que se refiera al mismo objeto sin alterar el valor de verdad del enunciado. Así, podemos reemplazar el término «212 grados Fahrenheit» por el término «100 grados centígrados» en la oración «El agua ordinaria hervirá al nivel del mar a 100 grados centígrados» y tendremos todavía una oración verdadera. [67-68]

Muchas oraciones que describen los estados mentales de las personas no satisfacen estas dos condiciones. La oración

> «Cathy cree que al nivel del mar el agua hierve a 100 grados centígrados»

contiene la oración:

> «el agua hierve a 100 grados centígrados»

pero el valor de verdad de la oración total no es una función del valor de verdad de este componente. La verdad del enunciado componente no nos informa de si el enunciado total es verdadero. Además, es fácil ver que carece de la condición de extensionalidad puesto que este enunciado de creencia puede ser verdadero y, con todo, la oración

> «Cathy cree que al nivel del mar el agua hierve a 212 grados Fahrenheit»

puede ser falsa. Si Cathy no sabe que 100 grados centígrados equivalen a 212 grados Fahrenheit, entonces ella no tiene creencias sobre la segunda oración. Si ella cree falsamente que 100 grados centígrados son equivalentes a 312 grados Fahrenheit, puede creer que es falso que el agua hierve a 212 grados Fahrenheit. (A este rasgo de los enunciados sobre los estados mentales se hace referencia como el *fallo de sustitutividad* [6].)

Muchos filósofos se refieren a las oraciones que exhiben esos rasgos lógicos como *oraciones intencionales*. Para llamar la atención sobre el hecho de que se diferencian en términos de esas peculiaridades lógicas (que las distingue de los enunciados extensionales), algunos filósofos usan la denominación *intensional* para esas oraciones [7]. [68-69]

Basándose en anomalías lógicas como las que se acaban de dar, Chisholm (1957,1958) intentó reformular la concepción de Brentano de la intencionalidad centrándose en los rasgos lógicos del lenguaje que usamos cuando hablamos sobre actividades psicológicas:

> Digamos: 1) que no necesitamos usar lenguaje intencional cuando describimos fenómenos no-psicológicos o «físicos»; podemos expresar todo lo que sabemos, o creemos, sobre tales fenómenos en un lenguaje que no es intencional. Y digamos: 2) que, cuando deseamos describir ciertos fenómenos psicológicos —en particular cuando deseamos describir pensar, creer, percibir, ver, conocer, desear, esperar y otros semejantes— o bien *a*) tenemos que usar un lenguaje que es intencional, o *b*) tenemos que usar un vocabulario que no necesitamos usar cuando describimos fenómenos no-psicológicos o «físicos» [8]. (1958, pp. 511-512.)

Chisholm mantiene que este modo de enmarcar el problema ofrece beneficios de los que carece la formulación original de Brentano. Evita el plantear el problema del status ontológico de los objetos intencionales limitando el foco de atención al lenguaje. Con todo, mantiene una distinción entre diferentes géneros de fenómenos en la naturaleza [9].

Se han planteado variadas objeciones al intento de Chisholm de caracterizar la intencionalidad lingüísticamente. Una de esas objeciones mantiene que tales criterios no cubren todas las oraciones sobre fenómenos mentales. Algunas oraciones como «Juan tiene dolor» o «Cathy está pensando en Carol» son claramente sobre fenómenos mentales pero no caen bajo ninguna de las tres condiciones de Chisholm (Cornman, 1962; Margolis, 1977). Otra objeción es que algunas oraciones que no son sobre fenómenos intencionales o psicológicos reúnen también las condiciones de Chisholm. Cualquier oración sobre lo que es posible o necesario, por ejemplo, mostrará [69-70] también fallo de sustitutividad. Tomando prestado un ejemplo de Quine (1953/1961b), es verdadero que

Nueve es el número de los planetas.

Es también verdadero que

Es necesario que nueve es mayor que siete.

Pero, si sustituimos los términos correferenciales, generamos el enunciado falso:

Es necesario que el número de los planetas es mayor que siete.

Se han hecho varios intentos de resolver estas dificultades y desarrollar un criterio lingüístico adecuado de intencionalidad (véase Chisholm, 1967; Lycan, 1969), y muchos filósofos aún aluden a tal criterio (ver Dennett , 1982; Rosenberg, 1980). Sin embargo, se han avanzado algunos poderosos argumentos en contra de proseguir con esta estrategia. Searle (1981), por ejemplo, argumenta que las peculiaridades lógicas que se encuentran en el lenguaje que describe los fenómenos mentales no caracterizan realmente rasgos del estado mental, sino sólo un rasgo del lenguaje usado para discutir los estados mentales. La intencionalidad se refiere al hecho de que los estados mentales tienen contenidos y que se refieren a otros fenómenos; y éstos son rasgos completamente diferentes del mundo más que peculiaridades lógicas de las oraciones sobre fenómenos mentales. Así pues, Searle pretende que la búsqueda de un criterio lingüístico es una quimera puesto que no se encamina a buscar los aspectos cruciales de la intencionalidad. (Para argumentos adicionales en contra de proseguir la búsqueda de un criterio lingüístico, ver Richardson, 1981.) Si se rechaza el criterio lingüístico, entonces uno parece verse forzado a volver a un criterio semejante al de Brentano y a la necesidad de hacer frente a la cuestión del status de los objetos intencionales.

3.4 LA REPRESENTACIÓN DE LOS ESTADOS INTENCIONALES COMO ACTITUDES PROPOSICIONALES

Otro enfoque para caracterizar la intencionalidad ha tomado pie a partir de la forma lingüística común de las oraciones que usan [70-71] verbos como «creer», «esperar», «desear» y otros semejantes. Los enunciados que usan esos verbos toman comúnmente la forma:

Cathy espera que su película reciba buenas críticas.

En esta forma, el verbo principal está seguido por la palabra «que» y una proposición. El verbo sirve para expresar la actitud de una persona hacia la proposición. Ésta es la razón por la que Russell (1940) introdujo la expresión «actitudes preposicionales» para referirse a tales oraciones. Esta forma se ha convertido en canónica a la hora de representar estados mentales. Aunque algunas veces usamos verbos como «esperar» y «creer» sin una proposición (como en «Jim cree a Cathy») tales oraciones pueden siempre transformarse en la forma canónica usando la palabra «que» al proporcionar una proposición

(por ejemplo, «Jim cree que lo que Cathy dijo es verdad»).

El formato canónico de la actitud preposicional es atractivo puesto que nos proporciona dos grados de libertad para caracterizar los estados mentales, representados por el verbo y la proposición. Puedes tener la misma actitud hacia diferentes proposiciones o diferentes actitudes hacia la misma proposición. Por ejemplo, puedes a la vez creer que Elena obtendrá el cargo y desear que no lo obtenga. Esta parece ser justamente la estructura correcta para explicar las acciones de una persona y hacer comparaciones entre los estados mentales de la gente. En primer lugar, la actitud y el deseo, cuando se ponen juntos y se dirigen hacia la misma proposición, pueden ser la causa de una acción (p. ej., diseñar un sabotaje para la candidatura de Elena. En segundo lugar, interpersonalmente, podemos dar cuenta de la diferencia entre dos acciones individuales poniendo de manifiesto cómo difieren en algunas de esas actitudes. Por ejemplo, dos personas pueden creer que es probable que Elena obtenga el cargo, pero una desea que lo obtenga, mientras que la otra puede desear que no.

Además de proporcionar un modo útil de caracterizar los estados mentales, la armazón de las actitudes preposicionales sugiere también un modo de caracterizar la intencionalidad de los estados mentales: usamos la proposición hacia la que la persona tiene una actitud para identificar el contenido del estado mental de la persona. El uso de proposiciones para especificar el contenido de estados mentales sugiere una conexión entre los análisis del lenguaje y de la mente. Esta conexión ha sido explotada por un gran número de filósofos, de modo que necesitamos considerar brevemente qué son las proposiciones. A menudo éstas se invocan en filosofía del lenguaje para representar el significado que podría ser compartido por diferentes [71-72] oraciones (p. ej., oraciones en lenguajes diferentes; ver p. 39). Como tal, una proposición se interpreta típicamente como una entidad abstracta, diferenciada, por una parte, de una oración particular emitida o escrita en un lenguaje y, por otra parte, del estado mental que llevó a alguien a emitirla o a escribirla. Se dice que una persona tiene una proposición en la mente cuando emite una oración, pero la proposición misma es algo separado del habla que el hablante capta o entiende. Los que invocan proposiciones al analizar el lenguaje las contemplan también como llevando a cabo otras funciones tales como servir de portadores de los valores de verdad («la proposición que Juan expresó era verdadera») y como singularizando el estado de cosas actual o posible al que se hace referencia en la oración [10].

Cuando se invocan así las proposiciones para explicar el discurso de las actitudes preposicionales usado para caracterizar estados mentales, éstas nos capacitan para explicar una ambigüedad importante. Cuando tanto tú como yo creemos que cenaremos en casa esta noche, nuestras creencias se dirigen a la proposición «Yo cenaré en casa esta noche». ¿Tenemos la misma creencia cuando compartimos la misma actitud proposicional hacia esta proposición? En un aspecto, la respuesta correcta parece ser sí y, en otro aspecto, no. La ambigüedad surge del hecho de que «casa» puede referirse a algún lugar particular (p. ej., mi casa), o a cualquier cosa que cuente como casa para la persona que cree. Cuando se refiere a cualquier cosa que cuente como casa, capturamos el aspecto en el que tanto tú como yo creemos la misma cosa cuando cada uno de nosotros cree que cenaremos en casa esta noche. Con todo, hay un aspecto en el que creemos algo completamente diferente, pues yo creo que cenaré en mi casa de Atlanta, mientras que tú crees que cenarás en un hogar distinto, probablemente en una ciudad diferente. De esta lectura da cuenta el hecho de que considero que «casa» se refiere a mi casa, mientras que tú consideras que se refiere a la tuya. Invocando la distinción de Frege entre el *sentido* y el *referente* de un término, en el

primer caso lo que importaba era el sentido de «casa», mientras que en el segundo era el referente. Esta distinción sentido-referencia, desarrollada en el análisis del lenguaje de Frege, nos permite entonces explicar la [72-73] ambigüedad que surge en las caracterizaciones de actitud proposicional de los estados mentales. (Para una discusión adicional, ver Dennett, 1982; Perry, 1977.)

La armazón de las proposiciones y de las actitudes preposicionales proporciona entonces un modo conveniente de caracterizar los estados mentales. Sirve también para localizar el problema de la intencionalidad, puesto que el propósito de citar la proposición es especificar el contenido del estado mental de alguien. Como veremos en el capítulo 4, la Teoría Computacional de la Mente intenta capitalizar esas ventajas. Hay, sin embargo, un peligro serio que surge cuando usamos formas de actitud proposicional para representar estados intencionales. Esta forma parece ofrecer una explicación de cómo surge la intencionalidad, pero no es así. La forma de actitud proposicional sugiere que el objeto de la actitud proposicional es la proposición misma, de modo que, por ejemplo, la creencia que uno tiene es *sobre* la proposición. Este movimiento encuentra el mismo problema que señalé al discutir el intento de Meinong de postular objetos intencionales como objetos de los estados mentales. El problema es que, si tratamos las proposiciones como los objetos de las actitudes intencionales, entonces todos nuestros estados mentales son sobre esas proposiciones y no *sobre* los objetos del mundo. Sin embargo, la intencionalidad de los estados mentales es justamente su capacidad de ser *sobre* eventos del mundo. Cuando invocamos las formas de actitud proposicional, tenemos que tener cuidado de recordar que las proposiciones han de ser las *portadoras* de la intencionalidad, no los objetos sobre los que son los estados intencionales. Al adscribir una actitud proposicional tal como

Sam cree que el gato es un animal feroz

la proposición

el gato es un animal feroz

enuncia lo que se cree, pero la creencia es *sobre* el gato y su putativa ferocidad, no meramente la proposición. Esto no va contra los intentos de algunos científicos cognitivos de usar los recursos de la estructura de actitud proposicional para desarrollar explicaciones del procesamiento mental. Significa, sin embargo, que el trabajo crítico de explicar la intencionalidad no se hace postulando la proposición o representación. La tarea de explicar cómo las proposiciones o las representaciones son *sobre* objetos o eventos del mundo, algunos de los cuales no existen de modo efectivo, queda por realizar.[73-74]

3.5 EL INTENTO DE NEGAR LA REALIDAD DE LA INTENCIONALIDAD

El uso de la actitud preposicional para representar estados mentales ha llevado también a filósofos como Quine, que cuestiona la legitimidad de las proposiciones como herramientas en el análisis del lenguaje, a cuestionar al mismo tiempo si la intencionalidad es un fenómeno real que nuestra ciencia deba intentar explicar. Los argumentos de Quine contra las proposiciones se basan generalmente en su tesis de

la indeterminación de la traducción. Esta tesis, discutida en el capítulo 2, mantiene que no hay significado determinado para los términos de un lenguaje puesto que siempre podemos establecer manuales alternativos para traducir términos del Lenguaje 1 al Lenguaje 2. Esos manuales alternativos harán equivaler los mismos términos del primer lenguaje con términos diferentes del segundo. No hay evidencia alguna, de acuerdo con Quine, que pueda mostrarnos que una de las traducciones es correcta. Quine contempló este argumento como si probase que es un error postular proposiciones determinadas para representar el significado de una oración puesto que la posibilidad de traducciones alternativas muestra que no hay un único significado. Además, él afirmó que es un error suponer que los hablantes tienen significados definidos en la mente cuando intentan emitir oraciones, puesto que nada nos impide emplear una traducción diferente y hacer, por tanto, una asignación diferente de significado.

Quine mantiene que es el punto de vista erróneo de que hay proposiciones el que da como resultado un punto de vista mentalista sobre el significado al que él se refiere como «el mito del museo» (Quine, 1969c). Este mito mantiene que hay estados mentales específicos, por ejemplo ideas o pensamientos, que podemos expresar cuando usamos el lenguaje. Quine afirma que se trata de un error, puesto que, lo mismo que podemos traducir oraciones de otro lenguaje de manera diferente dependiendo de qué manual de traducción elijamos, podemos interpretar la oración que usamos para especificar el contenido de una actitud preposicional de manera diferente dependiendo de qué manual de traducción elijamos. (La interpretación es, para Quine, comparable lógicamente a la traducción. En ambos casos estamos haciendo equivaler un conjunto de palabras con otro.) Imagínese que alguien intenta decirnos que él o ella cree que la evolución ha ocurrido por selección natural. Puesto que Quine afirma que podemos dar interpretaciones alternativas, en nuestras palabras, de la oración que representa lo que se cree, niega que haya nada determinado que la persona cree. Puesto que podemos aplicar [74-75] la tesis de la indeterminación a nuestro discurso interno traduciendo nuestras propias palabras a diferentes palabras de nuestro lenguaje, Quine niega también que haya algo determinado que nosotros creamos.

Quine considera esta tesis de la indeterminación como una muestra del error de pensar que la gente tiene estados mentales que exhiben intencionalidad. De hecho, él relaciona explícitamente su tesis de la indeterminación con la tesis de Brentano de que los estados mentales se caracterizan por la intencionalidad, pero extrae la conclusión opuesta de la de Brentano. Aunque Brentano mantuvo que tenemos que reconocer un *status* especial para los fenómenos mentales, Quine (1960) afirma que tenemos que expurgar nuestra ciencia de términos intencionales como *creencia*, incluyendo nuestra ciencia de la conducta humana:

> La tesis de Brentano de la irreductibilidad de los giros intencionales forma bloque con la tesis de la indeterminación de la traducción.
>
> Uno puede aceptar la tesis de Brentano o bien como mostrando la indespensabilidad de los giros intencionales y la importancia de una ciencia autónoma de lo intencional, o como mostrando la carencia de base de los giros intencionales y la vacuidad de una ciencia de la intención. Mi actitud, a diferencia de la de Brentano, es la segunda. Aceptar la usanza intencional en su valor facial es, hemos visto, postular la traducción relativa en principio a la totalidad de las disposiciones de habla. Tal postulación promete poca ganancia en penetración científica si no hay

para ella mejor fundamento que las supuestas relaciones de traducción que presupone el habla corriente de la semántica y de la intención (p. 221).

En lugar de una ciencia de la intencionalidad, Quine propone el desarrollo de un análisis conductista hasta sus últimas consecuencias de la conducta humana. Reconoce que usamos giros intencionales como *cree* en la vida diaria para describimos a nosotros mismos y a otros, pero, puesto que tales términos carecen de fundamento, tienen que eliminarse cuando nos volvemos a la ciencia: «Si estamos iluminando la verdadera y última estructura de la realidad, el esquema canónico para nosotros es el esquema austero que no conoce el estilo indirecto, sino el directo, y que no conoce las actitudes preposicionales, sino solamente la constitución física y la conducta de los organismos» (1960, p. 221).

El ataque de Quine a la noción de significado ha sido aceptado, con modificaciones, por un gran número de filósofos. Donald Davidson (1974a), por ejemplo, mantiene que cualquier adscripción de contenido a los enunciados o estados mentales de otras personas es un asunto de interpretación, no de descubrimiento. Putnam (1983) extrae una moraleja similar, manteniendo que la interpretación del [75-76] lenguaje o el pensamiento de otro es esencialmente una empresa holista llevada a cabo por un agente que interpreta. No es un asunto consistente en descubrir algo que pasa en la persona [11].

Otros, sin embargo, han hecho frente a las conclusiones de Quine. Algunos han desafiado la explicación de Quine del significado de la tesis de la indeterminación misma arguyendo que la decisión de adoptar un manual de traducción determinado y de desarrollar una teoría del significado para un lenguaje no es diferente de la decisión de aceptar una teoría particular en una disciplina científica y trabajar dentro de ella. Incluso si, como mantiene Quine, hubiese otras teorías empíricamente equivalentes a la que usamos, él concede que en física estamos autorizados a aceptar una teoría y a trabajar dentro de ella. Si tratamos las actividades de traducción e interpretación de una manera similar a la teorización en física, entonces contemplaremos la postulación de estados mentales para dar cuenta de los fenómenos intencionales como algo paralelo al desarrollar una teoría en física. La medida de adecuación de una teoría mentalista será el ver si sirve para nuestros propósitos científicos (p. ej., para explicar la conducta). Si sucede que el tratar a los seres humanos como teniendo estados intencionales facilita esos fines, entonces el favorecer tales estados estará perfectamente de acuerdo con el adoptar una actitud científica (ver Bechtel, 1978; Chomsky, 1969).

Quine, sin embargo, ha opuesto resistencia de forma resuelta a este enfoque argumentando que la tesis de la indeterminación establece algo más que el que las teorías mentalistas manifiestan la subdeterminación usual verdadera de todas las teorías científicas (Quine, 1969b). Él afirma que tales teorías son simplemente vacías. Sin embargo, si esas teorías son vacías es algo que parecería depender de su poder explicativo. Aunque el veredicto final no se ha [76-77] pronunciado todavía, el éxito de las teorías mentalistas que se han desarrollado en la ciencia cognitiva y las correspondientes limitaciones de los enfoques conductistas (Brewer, 1974) parecerían ser evidencia de que esas teorías tienen poderes explicativos de la misma manera que otras teorías científicas, y de este modo deberían ser tratadas de la misma manera que ellas (ver McCauley, 1987a; Palmer y Kimchi, 1986). Estamos obligados, pues, a explicar cómo surge la intencionalidad de esos estados mentales. En el capítulo 4 considero varias teorías que los filósofos han avanzado para dar cuenta de la intencionalidad de los estados mentales.

3.6 CONCLUSIONES PRELIMINARES SOBRE LA INTENCIONALIDAD

En este capítulo he introducido aquello a lo que los filósofos se refieren como la *intencionalidad* de los estados mentales: su capacidad de ser sobre cosas del mundo. También he examinado dos puntos de vista sobre cómo este rasgo de los estados mentales parece distinguirlos de otros estados puramente físicos. He mostrado también cómo podemos capturar la intencionalidad de los estados mentales describiéndolos en términos de actitudes preposicionales, donde las proposiciones enuncian el contenido de los estados mentales. Pero, con todo, esto no soluciona el problema de la intencionalidad, puesto que todavía tenemos que mostrar cómo se relacionan las proposiciones con los estados del mundo *sobre* los que se dice que son los estados mentales. Un enfoque de este problema es negar simplemente que existan los estados mentales intencionales. Así, Quine ha intentado negar la realidad de la intencionalidad y ha intentado mostrar que deberíamos limitarnos a una psicología conductista que no favorezca estados mentales. La ciencia cognitiva parece estar en el proceso de desarrollar teorías poderosamente explicativas que postulan estados mentales intencionales. Así parece que estamos frente al desafío de ver si no podemos explicar la intencionalidad de los estados mentales. En el capítulo 4 describo varias estrategias que los filósofos han usado para hacer justamente eso.

4. ESTRATEGIAS FILOSÓFICAS PARA EXPLICAR LA INTENCIONALIDAD

4.1 INTRODUCCIÓN

En el capítulo 3 he discutido varias concepciones de lo que es la intencionalidad y de cómo se piensa que señala una distinción entre fenómenos mentales y no mentales. Vimos cómo algunos filósofos, como Brentano, contemplaban la intencionalidad como algo que creaba un hiato entre los fenómenos mentales y los no mentales que prohibía el desarrollo de una ciencia de los fenómenos mentales comparable a las ciencias de los fenómenos puramente físicos. Vimos también cómo otros filósofos, como Quine, rechazaban la realidad de los fenómenos intencionales y proponían que la psicología no se centrase en absoluto en los fenómenos mentales, sino estrictamente en la conducta de los humanos y de otros organismos. La mayor parte de los científicos cognitivos encuentra ambas posiciones inadecuadas. En este capítulo describo algunas otras posiciones filosóficas que consideran que la intencionalidad es un rasgo real de los fenómenos mentales pero que intentan explicar cómo una ciencia que forme un continuo con las ciencias físicas puede dar cuenta de la intencionalidad.

4.2 LA TEORÍA COMPUTACIONAL DE LA MENTE (COMPUTACIONALISMO DE ÉLITE)

El primer enfoque que considero contempla la armazón de la actitud proposicional que usamos para describir los estados mentales de las personas como la base para una explicación científica de cómo opera de hecho la mente. En lugar de repudiar las proposiciones, este enfoque las trata como estructuras de la mente que sirven como el contenido de las actitudes mentales de una persona. El interés contemporáneo en este punto de vista ha sido inspirado por el desarrollo de los computadores. De acuerdo con una interpretación, puede pensarse en las proposiciones como símbolos en un computador digital moderno y en las actitudes hacia esas proposiciones como los modos en los que las configuraciones de esos símbolos se almacenan [78-79] en la memoria del computador. Por ejemplo, almacenar el símbolo o símbolos correspondientes a la proposición de que está lloviendo en el «cajón de creencia» constituiría la actitud proposicional de creer que está lloviendo. Este enfoque se extiende de los computadores a los humanos tratando a la mente como un computador que procesa símbolos en el que los símbolos se almacenan y se manipulan. Jerry Fodor (1980) se refirió a este punto de vista como la «Teoría Computacional de la Mente», mientras que Daniel Dennett (1986) la ha denominado «computacionalismo de élite».

Fodor ha sido el principal proponente contemporáneo de la Teoría Computacional de la Mente[1], cuya afirmación principal es que la psicología se ocupa de la estructura formal de los símbolos de la mente y del modo en que se manipulan. Puesto que los símbolos asumen el papel de proposiciones en el discurso de actitudes preposicionales y, de este modo, sirven para representar los fenómenos sobre los que uno está pensando, se denominan comúnmente representaciones mentales. Fodor propuso que la

mente poseía un conjunto de reglas que determinaban qué operaciones se realizan con esas representaciones. Esas reglas corresponden a los modos de inferencia que atribuimos a la gente en el discurso de actitudes proposicionales. Así, mientras que nosotros describiríamos a alguien como infiriendo la proposición «la comida campestre ha sido suspendida» a partir de la proposición «está lloviendo», la Teoría Computacional postula manipulaciones formales de símbolos representacionales (p. ej., moverlos dentro de diversos registros). Dados los papeles que desempeñan las reglas y las representaciones en tales explicaciones computacionales, a esas explicaciones se hace referencia algunas veces como «explicaciones de reglas-y-representaciones».

Fodor (1975) habló de esas representaciones mentales como constituyendo «un lenguaje del pensamiento». Añadió que la psicología sólo puede explicar la conducta humana si supone que los humanos razonan utilizando tal lenguaje interno. Para defender esta afirmación, Fodor señaló tres géneros de fenómenos. El primero era [79-80] la conducta racional. Cualquier explicación de la conducta racional debe tener en cuenta el que los organismos consideren las consecuencias de las acciones que están contemplando. Esto exige «que los agentes tengan medios para representar sus conductas a sí mismos: de hecho, medios para representar sus conductas en tanto que teniendo ciertas propiedades y no teniendo otras» (1975, p. 30). Por ejemplo, sólo si me represento a mí mismo que una consecuencia de no pagar mis impuestos es que iré a la cárcel seré capaz de tomar en consideración esta consecuencia a la hora de decidir si pago mis impuestos. El segundo fenómeno que Fodor consideró fue el concepto de aprendizaje. Fodor argumentó que solo podríamos aprender un nuevo concepto proponiendo una hipótesis sobre lo que el concepto podría significar y probando seguidamente su adecuación[2]. Por ejemplo, aprendemos el concepto «coche» haciendo la hipótesis de que se refiere a los objetos que cumplen ciertas especificaciones y, a continuación, probando si, de hecho, todos los objetos que cumplen esas especificaciones se cuentan como coches. Esto exige que poseamos de antemano un medio lingüístico en el que podamos enunciar tales hipótesis (ver Churchland, 1986, p. 389 para una refutación). El último fenómeno que Fodor señaló fue la percepción. De acuerdo con la tradición empirista, consideró la percepción como una actividad de resolución de problemas en la que el perceptor tenía que determinar lo que él o ella estaba viendo sobre la base de *inputs* sensoriales limitados. La percepción, al igual que el aprendizaje de conceptos, exige que el perceptor compruebe hipótesis (Fodor, 1975, p. 44). Tenemos que avanzar una hipótesis sobre lo que estarnos viendo (p. ej., que esto es un perro) antes de que podamos evaluar la evidencia a favor o en contra de la hipótesis.

Todos esos argumentos señalan, de acuerdo con Fodor, hacia la conclusión de que los agentes cognitivos tienen un sistema similar al lenguaje en el que llevar a cabo las actividades cognitivas. Un lenguaje natural ordinario como el castellano podría parecer un candidato para ser este sistema de lenguaje, pero Fodor mantuvo que no sería satisfactorio. En su lugar propuso que el lenguaje del pensamiento es un lenguaje interno, innato, que denominó «mentalés». Fodor ha ofrecido toda una serie de argumentos distintos a favor del mentalés. En primer lugar, los organismos que carecen de un lenguaje natural [80-81] pueden con todo realizar muchas de las actividades cognitivas que se acaban de describir. Por lo menos debe suponerse que tienen un lenguaje interno para manipular representaciones. (Patricia Churchland, 1978, respondió que esto reduce al absurdo la posición de Fodor.) En segundo lugar, el aprender un lenguaje exige un proceso de formación y puesta a prueba de hipótesis. Por lo menos, las hipótesis iniciales sobre el lenguaje natural no pueden representarse ellas mismas en el todavía-no-

conocido lenguaje natural y, de este modo, tienen que representarse en un lenguaje más básico[3].

Fodor contempló el proceso de pensar usando un lenguaje del pensamiento como algo que incluye solamente procesamiento sintáctico. La mente manipula símbolos sin consideración alguna hacia lo que se representa en esos símbolos. Esto lleva a Fodor a abrazar un punto de vista que Putnam (1975b) llamó «solipsismo metodológico»: el punto de vista de que desde la perspectiva de la psicología mentalista lo que está en el mundo no importa. Para Putnam el solipsismo metodológico revelaba la incompatibilidad de la psicología de las actitudes preposicionales y los enfoques computacionales de la psicología. Para mostrar esta incompatibilidad cuenta una historieta de ciencia ficción sobre un mundo posible, la Tierra Gemela, que es exactamente igual a nuestro planeta excepto en una cosa. En lugar de agua tiene otra substancia, XYZ, que se comporta igual que el agua y es indistinguible de ella. En la Tierra Gemela cada uno de nosotros tiene un duplicado, un *Doppelgänger*, que es idéntico a nosotros en todos los aspectos excepto en que él o ella tiene moléculas de XYZ en todas las partes donde nosotros tenemos moléculas de H_2O. Puesto que somos iguales en todos los aspectos, se sigue que mi *Doppelgänger* y yo tenemos que tener los mismos estados psicológicos. En particular, ambos afirmamos la oración «Estoy bebiendo agua». A pesar del hecho de que mi *Doppelgänger* y yo estemos en el mismo estado psicológico, con todo queremos decir cosas diferentes mediante esas palabras. Mi enunciado es sobre H_2O, mientras que el de mi *Doppelgänger* es sobre XYZ. La moraleja que Putnam extrae de esta historieta es que los significados no están en la cabeza: lo que determina el referente de mi término *agua* no depende solamente de mi [81-82] estado psicológico, sino también de las cosas con las que yo estoy conectado causalmente. Puesto que una de las funciones clásicas de las proposiciones era proporcionar el significado de las oraciones y determinar sus extensiones, Putnam afirma que las representaciones que se considera que están en la cabeza por parte de la explicación computacional de la psicología, no son lo mismo que las proposiciones de la psicología de las actitudes preposicionales. (Ver Burge, 1979, 1982; Stich, 1978, 1983 para argumentos relacionados.)

Para Putnam los enfoques computacionales de la psicología son solipsistas en la medida en que no pueden habérselas con ese aspecto del significado que depende del mundo. Putnam consideró esto como una desventaja, pero Fodor (1980) extrajo una moraleja distinta. El enfoque apropiado para la psicología es, de acuerdo con Fodor, emplear las mismas proposiciones que figuran en la psicología de las actitudes preposicionales para desarrollar una explicación de lo que sucede en la mente. Si algo del significado de esas proposiciones se pierde al tratarlas como estructuras en la cabeza, entonces la psicología debe conformarse con las estructuras sintácticas que podrían estar en la cabeza. En defensa de este punto de vista él afirma en primer lugar que la única cosa que puede influir en nuestra conducta es lo que está representado formalmente dentro del sistema. El si existimos en un mundo de H_2O o de XYZ no afecta a nuestra conducta a menos de que afecte a nuestras estructuras internas: «Lo que el agente tiene en su mente es lo que causa su conducta», no aquello a lo que esos estados mentales se refieren (Fodor, 1980, p. 67). En segundo lugar, Fodor afirma que es una suerte que la psicología se limite a usar esas estructuras formales al explicar la conducta, puesto que, de lo contrario, tendríamos que descubrir conexiones legaliformes entre representaciones y objetos externos. Pero ésas no son posibles a menos que podamos identificar los géneros naturales correctos que sirven como referentes de nuestras representaciones mentales[4]. Sólo poseeremos tal conocimiento una vez que todas las demás ciencias hayan completado su trabajo y hayan descubierto los verdaderos géneros naturales[5]. (Ver Field, 1978,

para argumentos adicionales a favor de la Teoría Computacional.) [82-83]

Uno de los rasgos atractivos de la Teoría Computacional es que puede explicar fácilmente peculiaridades lógicas del discurso sobre los estados mentales tales como el fallo de substitutividad de expresiones correferenciales (ver capítulo 3). La Teoría Computacional mantiene que el sistema cognitivo puede sólo realizar aquellas manipulaciones sancionadas por las reglas y representaciones que tiene. Considérese cómo podría funcionar una explicación computacional de Edipo. Al principio de la obra Edipo rey, Edipo aprende que Yocasta es la reina y quiere casarse con ella. Pero, aunque Edipo no lo sabe, Yocasta es también su madre. En el modelo computacional Edipo almacenaría la proposición

Estoy casado con Yocasta

en su cajón de creencia. El modelo posee una regla que le permite substituir un nombre por otro cuando son correferenciales. Pero en este estadio el sistema no sabe que «Yocasta» y «mi madre» son correferenciales y no lleva a cabo la substitución. Cuando Edipo, al avanzar la obra, aprende esta información, ésta se representa formalmente en el modelo. Ahora, de un modo puramente formal, el modelo infiere la nueva oración

Estoy casado con mi madre.

Aunque la Teoría Computacional puede explicar así el fallo de substitutividad de las expresiones correferenciales en descripciones de los estados mentales de Edipo, no encara tan claramente el problema de cómo esos estados mentales pueden ser *sobre* algo. Las representaciones que la Teoría Computacional atribuye a la mente tienen, se supone, una función referencial, pero la teoría no explica cómo realizan esa función. Así, Richardson (1981) objeta que la Teoría Computacional, al igual que cualquier teoría que postule objetos intencionales, simplemente pospone el problema de explicar la intencionalidad. Para explicar la intencionalidad de los estados mentales [83-84] tenemos que explicar cómo se conectan las representaciones con objetos del mundo. Si no podemos dar cuenta de esto, quedamos en una posición en la que tratamos las actividades de pensar como algo totalmente separado del mundo natural. Fodor (1987) ha desarrollado una estrategia alternativa para atacar este problema. Explica cómo las representaciones mentales son sobre rasgos del mundo en términos de sus conexiones causales con estados externos del mundo. Un enfoque como éste debe vencer, sin embargo, un serio obstáculo que mencionamos al principio. Una de las características que distinguen la intencionalidad de los estados mentales es que pueden deformar la situación real del mundo y ser sobre cosas que no existen. Una enfoque causal corre el riesgo de conectar todo estado mental con un estado extemo y hacer así imposible malrepresentar estados del mundo y referirse a entidades no existentes.

4.3 REPRESENTACIONES SIN COMPUTACIONES

La Teoría Computacional no es solamente una propuesta filosófica especulativa. Muchos investigadores de inteligencia artificial (IA) también contemplan los sistemas cognitivos como

manipuladores de símbolos formales. Ellos intentan desarrollar estructuras de símbolos formales que puedan producir conducta inteligente[6]. Sin embargo, numerosos filósofos han criticado la Teoría Computacional de la Mente o bien tal como la defiende Fodor o tal como figura en IA. Muchos de los que la rechazan aceptan el punto de vista de que los sistemas cognitivos representan cosas y que, por tanto, son sistemas intencionales. Lo que niegan es que esto exija estados específicos dentro de sistemas cognitivos que se empleen como representaciones y que sean manipulados por reglas formales. Todo lo que exige, mantienen ellos, es que haya actividad de algún tipo en el sistema que explique cómo éste tiene estados mentales que son sobre cosas. En esta sección describo varios argumentos en contra de la explicación computacional. Las propuestas específicas acerca de cómo la mente puede ser representacional, y por tanto intencional, sin realizar computaciones sobre representaciones, se discuten en secciones posteriores. [84-85]

La primera objeción contra la Teoría Computacional es que es empíricamente poco plausible en tanto que explicación de las capacidades cognitivas humanas. Dennett (1977) planteó esta objeción en su recensión del libro de Fodor *The Language of Thought*:

> Fodor parece suponer que las únicas estructuras que podrían garantizar y explicar el poder predictivo de nuestros cálculos intencionales tienen que reflejar la sintaxis de esos cálculos. Esto o es trivialmente verdadero (puesto que la estructura «sintáctica» de los eventos o estados se define simplemente por su función) o es una afirmación empírica que es muy interesante, no enteramente implausible y, con todo, ni demostrada ni siquiera argumentada por lo que yo sé hasta ahora. Por ejemplo, supongamos que los *hamsters* son interpretables como buenos bayesianos por lo que respecta a las decisiones que toman. ¿Tenemos en principio que ser capaces de encontrar algunos rasgos destacables en los controles de los *hamsters* que sean interpretables como instancias de fórmulas en algún cálculo bayesiano? Si ésta es la conclusión de Fodor, no veo que le haya dado el apoyo que necesita y confieso que no creo en ella en absoluto (p. 279).

Más recientemente Dennett (1986) ha afirmado que «un cerebro que manipulase símbolos computacionales parece profundamente no biológico» (p. 66). Igualmente escéptico es P. S. Churchland, que mantiene que «el *crunching* de las oraciones parece insensible a consideraciones evolucionistas» (1986, p. 138; ver también P. S. Churchland, 1980a). Ella plantea un dilema evolucionista al defensor de un lenguaje del pensamiento: o bien tenemos que contemplar el procesamiento de oraciones como algo que surge muy pronto en la filogenia, o tenemos que afirmar que los procedimientos de procesamiento de oraciones empleados en la cognición humana no tienen raíz alguna en los procesos mentales de otros organismos. La segunda opción es insatisfactoria, puesto que los humanos no lingüísticos, así como los miembros no lingüísticos de otras especies, parecen claramente ser capaces de planificación racional, y así parecen participar en el mismo tipo de acciones cognitivas en las que nosotros participamos. De otro lado, la suposición de que los organismos no lingüísticos y prelingüísticos que manifiestan cognición poseen todos un lenguaje del pensamiento completo la considera ella como algo fuertemente implausible (véase Kitcher, 1984, para una respuesta).

La explicación computacional de la cognición parece empíricamente problemática en otros aspectos. Puesto que el sistema ha de operar con reglas puramente formales o sintácticas para manipular

representaciones, todo aspecto del significado del símbolo que ha de afectar al procesamiento psicológico debe estar formalmente codificado. Trabajando totalmente de acuerdo con principios sintácticos, el sistema no tendrá acceso a los contextos que, en el lenguaje natural, sirven para eliminar ambigüedades en los diferentes significados [85-86] de los términos. Al analizar los lenguajes naturales, Searle (1979) ha argumentado que es inútil desarrollar explicaciones formales o sintácticas de los significados de expresiones, puesto que esas expresiones toman frecuentemente diferentes significados en contextos diferentes. Pero esto es exactamente lo que exige la Teoría Computacional. Esta objeción es de hecho anterior a las teorías computacionales modernas de Fodor y de la IA. Muy al principio de este siglo Husserl propuso una explicación de la cognición en términos de la manipulación de proposiciones almacenadas[7]. Martin Heidegger (1949/1962) se opuso al programa de Husserl sobre la base de que la variabilidad de la información con la que hemos de vérnoslas no podría expresarse adecuadamente por medio de proposiciones fijas. Heidegger propuso que el modo de vencer este problema es reconocer que alguna de la información que empleamos no está representada en la mente, sino que se encuentra en cosas tales como en las destrezas refinadas y en nuestro nexo social. Herbert Dreyfus (1979) ha desarrollado adicionalmente las objeciones de Heidegger en sus propias críticas de la IA y concluye que la IA está descaminada cuando intenta representar toda la información que usa un sistema cognitivo en términos de símbolos sintácticos almacenados en la cabeza (ver capítulo 7 para una discusión adicional de la posición de Dreyfus).

Una objeción adicional a la Teoría Computacional se concentra en el número de tales oraciones mentales que cada uno de nosotros debe poseer si la explicación es correcta. Si todo estado mental ha de entenderse como alguna forma de almacenamiento o procesamiento de una oración en el lenguaje del pensamiento, cada uno de nosotros necesitaría tener un número infinito de tales oraciones almacenadas en nuestra mente/cerebro. La razón es que tenemos un número infinito de creencias, muchas de las cuales jamás consideramos conscientemente de forma activa. Por ejemplo, muchos de nosotros creemos que las cebras no llevan abrigo, aunque es dudoso que muchos de nosotros hubiésemos considerado conscientemente esa proposición hasta que Dennett la introdujo como un ejemplo. Similarmente, muchos de nosotros creemos que los osos miden menos de *n* pies de alto para todo *n* mayor que siete (véase P. S. Churchland, 1986). Los críticos del punto de vista computacional afirman que tal conjunto infinito de oraciones mentales no podría almacenarse en la mente/ cerebro.[8] [86-87]

La afirmación de la Teoría Computacional de que todo el conocimiento ha de representarse sintácticamente genera todavía otros problemas. Uno de ellos tiene que ver con cómo identificamos la información relevante para una tarea particular. Aquellos que diseñan sistemas de inteligencia artificial se enfrentan ya con tales problemas respecto de sistemas que tienen relativamente poca información almacenada. Dennett (1984a) ilustró este problema, comúnmente conocido como el «problema del contorno» («*frame problem*») (McCarthy y Hayes, 1969), en términos de una historia en la que a un robot se le dice que su suministro de energía está en una habitación donde se ha puesto una bomba que va a estallar. El robot tiene que decidir cómo salvar su suministro de energía reuniendo la información relevante y haciendo las inferencias apropiadas. El problema es proporcionar el conjunto correcto de reglas para hacer esto. Tal robot tiene que responder a varias contingencias, cada una de las cuales hace relevante para su tarea a diferentes piezas de información. No puede buscar toda la información sin verse atrapado en un proceso de razonamiento sin fin. A medida que se almacena más información en las

representaciones formales, esta tarea se vuelve aún más difícil (véase también Dreyfus, 1985).

Una última objeción al punto de vista computacional se centra en la dificultad de determinar siquiera el carácter efectivo de las representaciones formales del lenguaje del pensamiento. Dennett (1982) afirmaba que, si postulamos un sistema de símbolos sintáctico tal como el del lenguaje del pensamiento de Fodor, deberíamos ser capaces de plantear la cuestión de si todos nosotros tenemos el mismo lenguaje del pensamiento o lenguajes diferentes. Las diferencias en nuestros lenguajes del pensamiento podrían explicar diferencias cognitivas, pero, puesto que no tenemos ningún modo independiente de identificar diferencias en nuestros lenguajes del pensamiento, tal explicación se convierte en circular. Además, sabemos por los lenguajes naturales que diferentes mensajes pueden transmitirse en el mismo lenguaje y que diferentes lenguajes pueden transmitir el mismo mensaje. De este modo, no podemos inferir similaridades o diferencias en el modo en que se usan. Nos quedamos postulando un [87-88] lenguaje del pensamiento sobre el que aparentemente no podemos decir nada.

Si encontramos que argumentos[9] tales como éstos son suficientes para rechazar la Teoría Computacional, nos queda el desafío de mostrar cómo un sistema cognitivo representa cosas y es así intencional. Una de las principales virtudes del enfoque computacional era que estaba diseñado para captar el modo en que ordinariamente describimos estados mentales en términos de actitudes preposicionales, y de este modo adquiría todos los beneficios de tal enfoque. Dennett afirmaba, sin embargo, que podemos emplear la armazón de la actitud proposicional para describir personas sin hacer equivaler las proposiciones con símbolos formales en la mente. Para hacer esto necesitamos una explicación alternativa de lo que está incluido cuando alguien «capta» una proposición. Dennett (1982) propuso lo siguiente: «Las proposiciones son captables si y sólo si los predicados de actitud proposicional son predicados proyectables, predictivos y disciplinados de una teoría psicológica» (p. 10). Todo lo que esto exige es que nuestras adscripciones teóricas de actitudes preposicionales covarían con predicciones sobre la conducta. No exige adicionalmente que lo que sucede dentro de la mente sea computación de proposiciones.

P. M. Churchland ha defendido también el usar actitudes preposicionales sin invocar la Teoría Computacional. Él ha comparado las predicciones hechas en el discurso de actitudes preposicionales con las predicaciones hechas en las ciencias físicas, muchas de las cuales no tienen entrenamientos ontológicos especiales:

> La ironía es que cuando examinamos la estructura lógica de nuestras concepciones populares en este punto, no encontramos diferencias, sino algunas *similaridades* muy profundas entre la estructura de la psicología popular y la estructura de las teorías físicas paradigmáticas. Comencemos comparando los elementos de las dos listas siguientes:
>
> | *Actitudes proposicionales* | *Actitudes numéricas* |
> |---|---|
> | ... cree que P | ... tiene una longitud $_{m}$ de *n* |
> | ... desea que P | ... tiene una velocidad $_{m/s}$ de *n* |
> | ... teme que P | ... tiene una temperatura $_{k}$ de *n* |
> | ... ve que P | ... tiene una carga $_{c}$ de *n* |

... sospecha que P ... tiene una energía cinética $_j$ de *n*

Donde la psicología popular exhibe actitudes *proposicionales*, la física matemática exhibe actitudes *numéricas*. (Churchland, 1984, p. 64.) [88-89]

Churchland mantuvo que podemos desarrollar leyes que se refieran a actitudes preposicionales justamente cuando podemos desarrollar leyes que se refieran a actitudes numéricas. Así como hablar de actitudes numéricas no nos compromete con la postulación de ninguna entidad especial —velocidad m/s —, tampoco el hablar de actitudes preposicionales nos compromete con tratar a las proposiciones como entidades[10].

Aunque Dennett y Churchland afirman que podemos emplear todavía el enfoque de la actitud proposicional sin suscribir la Teoría Computacional, Fodor puede con todo objetar que éstos no nos han dicho qué actividades en la cabeza de una persona capacitan a esa persona para representar su entorno. Fodor afirma que es una virtud de la Teoría Computacional el que sea capaz de hacer esto. Considérese la argumentación que él hace a favor de un lenguaje del pensamiento (Fodor, 1975):

1. Los únicos modelos psicológicos de procesos cognitivos que parecen siquiera remotamente plausibles representan tales procesos como computacionales.
2. La computación presupone un medio de computación: un sistema representacional.
3. Las teorías remotamente plausibles son mejores que ninguna teoría en absoluto.
4. Estamos entonces provisionalmente comprometidos en atribuir a los organismos un sistema representacional. (p. 27).

La tercera premisa del argumento de Fodor parece enteramente razonable, e impone una carga sobre cualquiera que se oponga a su conclusión. Uno debe o bien presentar modelos de cognición que no sean computacionales o mostrar que la computación no presupone un sistema representacional.

Stich (1983), al defender aquello a lo que él se refiere como la teoría sintáctica de la mente, ha rechazado la segunda premisa del argumento de Fodor argumentando que aunque las operaciones dentro de la mente pueden interpretarse como operaciones formales o sintácticas semejantes a aquellas de una teoría sintáctica (en lingüística), los objetos sobre los que se realizan esas operaciones sintácticas no necesitan contemplarse como representaciones, esto es: como [89-90] unidades a las que puede asignarse contenido. Stich afirmaba que gran parte del trabajo tanto en inteligencia artificial como en psicología cognitiva tenía este carácter. Los investigadores de esos campos postulan procedimientos sintácticos para explicar la conducta, pero no exigen que todos los objetos sintácticos usados para producir el *output* se interpreten como representando algo (véase Von Eckardt, 1984, para un argumento relacionado.) El enfoque de Stich es entonces computacional pero no mantiene que las entidades que vayan a ser manipuladas sean representaciones. Éste parece haber sido el enfoque de muchos investigadores en ciencia cognitiva a pesar de las afirmaciones de Fodor en sentido contrario (véase, sin embargo, McCauley, 1987).

Más recientemente, sin embargo, gran número de los que practican la ciencia cognitiva han propuesto un programa que rechaza la primera premisa de Fodor (que los únicos modelos psicológicos remotamente plausibles son computacionales). Los abogados de los modelos «conexionista» o de «procesamiento distribuido paralelo» (PDP) han propuesto modos de modelar los fenómenos cognitivos que no son

computacionales en el sentido usado aquí. Éstos no realizan operaciones sobre símbolos almacenados a la manera de un computador de von Neumann. Dicho brevemente, lo que esos investigadores están haciendo es explorar las capacidades de una clase de sistemas diseñados sobre el modelo de las redes neurales. Los sistemas constan de nudos, cada uno de los cuales tiene un grado determinado de activación en cualquier tiempo y se conecta con cierto número de otros nudos a los que envía estímulos inhibidores o excitadores. Cuando se da un modelo inicial de activación, las excitaciones e inhibiciones que pasan a través del sistema alterarán los estados de activación de los nudos hasta que se consigue un modelo estable. Las intensidades de las conexiones excitativas e inhibitivas pueden estar diseñadas de manera que cambien como resultado de la actividad local del sistema. Cuando los sistemas están diseñados de esta manera, pueden aprender a responder de nuevas maneras con el resultado de que se colocarán en estados diferentes en ocasiones posteriores. Lo que tiene interés es que los investigadores han empleado tales sistemas (tal como están simulados en computadores de von Neumann) para modelar ciertas funciones cognitivas. En tareas como la de reconocimiento de patrones sus ejecuciones son mucho más semejantes a las humanas que las de las máquinas que procesan reglas. En esas simulaciones, los investigadores interpretan la actividad del sistema y de este modo tratan al sistema como representacional, pero el sistema no opera realizando computaciones sobre representaciones. Los modelos conexionistas proporcionan un ejemplo de cómo es posible [90-91] desarrollar una teoría representacional sin una Teoría Computacional[11].

La llegada de los modelos conexionistas proporciona apoyo a aquellos que suscriben la Teoría Representacional de la Mente pero rechazan la Teoría Computacional. Con todo, la Teoría Representacional de la Mente no explica la intencionalidad porque no explica cómo es capaz la mente de representar cosas. Recientemente han surgido tres teorías filosóficas distintas, cada una de las cuales ha intentado explicar cómo la mente/cerebro puede representar cosas y ser así intencional: *a)* el enfoque de la teoría de la información, *b)* el enfoque de la reducción biológica, y *c)* el enfoque de la postura intencional. En las restantes cuatro secciones se procede a su discusión.

4.4 EL ENFOQUE DE LA TEORÍA DE LA INFORMACIÓN

Puesto que los estados intencionales son estados que llevan información sobre otros estados, algunos filósofos han buscado explicar la intencionalidad apelando a la teoría matemática de la información avanzada por Shannon y Weaver (1949). La apelación a la teoría matemática de la información se ha rechazado a menudo sobre la base de que la teoría de la información se ocupa de la capacidad de los canales para transportar información y no de la información particular que transportan. Fred Dretske (1981; ver también 1983 y los comentarios siguientes) ha argumentado, sin embargo, que hay una útil intuición en la teoría de la información que puede explotarse. Se trata de la idea de que un estado transporta información sobre otro justamente en el grado en que depende de acuerdo con leyes de ese otro estado. Dretske propuso que, si hay una relación determinista y de acuerdo con leyes de manera que yo puedo inferir de la señal que ésta [91-92] tuvo una causa particular, entonces la señal me da una información *sobre* la causa. La relación de acuerdo con leyes entre causa y señal da cuenta, mantiene él, de que la señal es *sobre* la causa. Entonces, *ser sobre* no es un rasgo único de los estados mentales, sino que se halla en todas las relaciones causales:

> *Cualquier* sistema físico cuyos estados internos son dependientes de acuerdo con leyes, en algún modo estadísticamente significativo, del valor de una magnitud externa (de la manera en que un instrumento de medida apropiadamente conectado es sensible al valor de la cantidad para cuya medición está diseñado) cumple los requisitos de un sistema intencional. (Dretske, 1980, p. 286.)

El desafío, como Dretske vio, no consiste en explicar cómo algo podría manifestar intencionalidad, sino explicar cómo algo podría exhibir el género correcto de intencionalidad para ser una mente. Lo que es característico de nuestros estados mentales no es que tengan contenido, sino los contenidos específicos que tienen. Un instrumento de medida típico conlleva generalmente mucha más información que nosotros, los usuarios, adquirimos de él. Conlleva información sobre cada paso de la etiología causal de la lectura del instrumento. Nuestros estados cognitivos distinguen entre diferentes contenidos que se registran de manera indiscriminada mediante los instrumentos de medida típicos. Los contenidos de los estados mentales son las propiedades medidas; no todos los estados intermedios causalmente necesarios. Para captar esta diferencia, Dretske (1983) distinguió entre lo que él llama información «digital» y «analógica» en la percepción:

> Al pasar de la representación sensorial a la cognitiva (de ver una manzana a darse cuenta de que se trata de una manzana), hay un despojar sistemático de componentes de la información (que se relacionan con el tamaño, el color, la orientación y el entorno), que hace que la experiencia de la manzana sea lo fenoménicamente rica que sabemos que es, para que se destaque *un* componente de esta información —la información de que es una manzana—. La digitalización (de, por ejemplo, la información de que *s* es una manzana) es un proceso mediante el cual un trozo de información se separa de una matriz de información más rica en la representación sensible (mientras que se mantiene en lo que yo llamo forma «análoga») y se destaca excluyendo a todo lo demás (p. 61).

Dretske le dio la vuelta al modo normal en que pensamos sobre la intencionalidad. Su análisis causal hace que casi todo estado sea intencional y así, más bien que preguntar cómo algunos estados llegan a tener la característica única de la intencionalidad, la tarea de Dretske es explicar cómo algún *status* tiene intencionalidad concentrada y [92-93] limitada. En su análisis Dretske subrayó el lado relacional de la intencionalidad (ver p. 43), y esto plantea la cuestión de si Dretske puede dar cuenta de los estados intencionales que no logran referirse a nada real. El problema puede reconocerse contemplando la explicación de Dretske de la manera que él pretende: como parte de un proyecto epistemológico diseñado para explicar lo que es el conocimiento y cómo es posible. Además, en epistemología Dretske le dio la vuelta a la estrategia normal, que es comenzar con la creencia y preguntar bajo qué condiciones una creencia cuenta como conocimiento. En la explicación de Dretske todos los estados informacionales conllevan automáticamente conocimiento; el desafío es mostrar cómo podríamos llegar a tener creencias falsas. Esto incluye mostrar cómo podría ir mal el proceso de extracción y, de esta manera, representar equivocadamente cosas del mundo.

Muchos comentadores de la explicación de Dretske afirman que su tratamiento de las creencias falsas y de los fallos de referencia se debe a distorsiones de conocimiento que, por otra parte, es verídico,

resultando descaminada la información referencial. Una implicación del enfoque de Dretske parecería ser que para poseer conocimiento necesitaríamos simplemente eliminar los errores inducidos por nuestro sistema cognitivo. Entonces podríamos recobrar el Edén en el que poseyéramos información incorrupta (ver Churchland y Churchland, 1983). Esto parece denigrar la mente contemplándola como un agente de distorsión tanto por lo que respecta al conocimiento, como por lo que respecta a la intencionalidad. Tal punto de vista casa mal con una perspectiva evolucionista que interpretaría a las mentes de los organismos más elevados como algo que mejoraría la capacidad del organismo para ganar información, no como algo que impusiese distorsión[12].

Aunque muchos filósofos (p. ej., Fodor, 1984) mantienen que Dretske ha enfocado la intencionalidad de la manera incorrecta, su enfoque tiene ciertamente un atractivo. Hace que el aspecto de la intencionalidad consistente en «ser sobre» sea totalmente natural en la medida en que emerge como un aspecto de las relaciones causales ordinarias [93-94] Este enfoque sería particularmente atractivo si no pareciese reducir los estados mentales a productos potencialmente distorsionados de estados de *input* fiables. John Heil (1983) desarrolló una explicación que es similar en algunos aspectos a la de Dretske, pero que introduce componentes cognitivos de una manera más constitutiva. Él está de acuerdo con Dretske y con el psicólogo J. J. Gibson (1979) al tratar la información como algo que está presente en nuestro entorno y está dispuesto para ser «seleccionado» por agentes cognitivos. Al igual que Dretske y Gibson, Heil trató la selección de información como algo que genera causalmente los estados mentales del que conoce. Sin embargo, se diferencia de ambos al caracterizar esos estados mentales de una manera neokantiana (capítulo 1), al mantener que los estados mentales surgen solamente una vez que la experiencia perceptiva se conceptualiza usando conceptos proporcionados por el agente. La posición, sin embargo, no es totalmente kantiana desde el momento en que insiste en que el proceso de conceptualización no es inferencial, sino causal. Dado tanto el aparato perceptivo del que conoce, como su armazón conceptual, la información del estímulo sensorial causa creencias en la persona. Así, Heil no está de acuerdo con Dretske cuando mantiene que la información del entorno no es intencional porque no está conceptualizada. Pero, con todo, también para Heil la información figura de manera central al explicar la intencionalidad de los estados mentales. Sirve para conectar estados del agente con rasgos de un entorno. En la última parte de este capítulo discuto cómo un punto de vista relacionado por lo que respecta a las relaciones entre el organismo y el entorno puede figurar en un análisis de la intencionalidad.

4.5 EL ENFOQUE DE LA REDUCCIÓN BIOLÓGICA

Aunque el análisis de Dretske intenta mostrar que la intencionalidad es, de manera general, un rasgo de la naturaleza, John Searle ha argumentado que se trata de un rasgo que se encuentra sólo en ciertos sistemas biológicos y, de este modo, exige una explicación biológica, no una explicación por parte de la ciencia cognitiva. Al defender un análisis biológico Searle no nos dice qué rasgos de los sistemas biológicos los hacen intencionales (de hecho, argumento más tarde que un rasgo peculiar de la posición de Searle es que no puede lógicamente intentar esto). Más bien, él mantiene simplemente que sólo una teoría biológica podría explicar la intencionalidad.

Searle caracteriza la intencionalidad inspirándose en el análisis de los actos de habla que él había

desarrollado previamente (Searle, 1969, 1979; ver capítulo 2 de este volumen).[94-95] La intencionalidad, tanto de los actos de habla como de los estados intencionales, consiste para Searle en lo que él llama su «direccionalidad de ajuste». Algunos actos de habla y estados mentales se supone que se corresponden con el modo en que es el mundo, mientras que otros imponen sobre el mundo la obligación de que se corresponda con ellos. Una creencia, por ejemplo, se supone que se corresponde con el modo en que es el mundo, mientras que una orden impone sobre el mundo la obligación de que se corresponda con ella. En muchos estados mentales hay una conexión causal, así como una relación semántica entre el estado intencional y el mundo. Así, pues, las experiencias perceptivas dependen causalmente de cosas del mundo mientras que las órdenes pueden causar ciertos efectos en el mundo. Searle (1981) retrató esas conexiones causales como dando la vuelta a las relaciones de dirección de ajuste:

> Las experiencias perceptivas tienen la dirección de ajuste mente-a-mundo y la dirección de causación mundo-a-mente (aproximadamente, esto significa que sólo se satisfacen si el mundo es como parece ser y si su ser de esa manera causa el que perceptualmente parezca de esa manera), mientras que las intenciones en la acción son exactamente opuestas tanto en dirección de ajuste como en dirección de causación. Tienen dirección de ajuste mundo-a-mente y dirección de causación mente-a-mundo (esto significa que se satisfacen solamente si el mundo llega a ser del modo en que uno intenta que sea y si su ser de esa manera viene causado por el hecho de que uno intente hacer que sea de esa manera) (p. 729).

Aunque ha tratado el análisis de los actos de habla como proporcionando un modelo útil para desarrollar él análisis de la intencionalidad, Searle ha insistido en que la intencionalidad de los estados mentales es más básica. El lenguaje no tiene lo que Searle ha denominado «intencionalidad intrínseca», sino solamente «intencionalidad derivada», algo que adquiere a partir del estado mental subyacente. Searle no ofrece una explicación positiva de lo que es la intencionalidad intrínseca, sino que se establece la tarea de mostrarnos aquello que carece de ella. Al igual que los actos de habla, Searle mantuvo que los computadores sólo tienen intencionalidad derivada. Él ha rechazado la Teoría Computacional de la Mente, puesto que mantiene que los procesos computacionales son insuficientes para dar cuenta de la intencionalidad intrínseca.

Searle ha argumentado a favor de esas afirmaciones presentando un *Gedankenexperiment* (experimento de pensamiento) en el que él se imaginaba a sí mismo desempeñando el papel de un computador que está programado para responder preguntas sobre una historieta [95-96] (aquí simplifico un poco)[13]. El elemento crucial en la explicación de Searle es que tanto la historieta, como las preguntas, como los *outputs* de Searle, están en chino, un lenguaje que él no entiende. Él es capaz de «responder» a las preguntas solamente porque, junto con los símbolos chinos que contienen la historieta y las preguntas, ha recibido reglas enunciadas en inglés que le dicen cómo producir nuevas ristras de símbolos dependiendo de las ristras que encuentra en la historieta y en las listas de preguntas. Todo el asunto está diseñado tan inteligentemente que, mientras que Searle creía que estaba sólo manipulando símbolos y no sabía que está respondiendo a preguntas en chino sobre una historieta en chino, de hecho estaba produciendo un *output* perfectamente coherente que los hablantes nativos del chino considerarían auténtico. Searle mantiene que, puesto que él no entendía lo que estaba haciendo, sus actividades de

manipulación no podrían considerarse como intencionales en el sentido de ser sobre aquello de lo que trata la historieta (como hubiera sucedido si la historieta y las preguntas se hubiesen presentado en inglés). Además, puesto que él estaba ejecutando todas las manipulaciones de símbolos formales que postula el análisis computacional y, con todo, no entendía, el análisis computacional es inadecuado:

> En el caso del chino tengo todo lo que la inteligencia artificial puede poner en mí por medio de un programa, y no entiendo nada, y no hay hasta ahora razón alguna para suponer que mi comprensión tenga algo que ver con los programas de computador, esto es: con operaciones computacionales realizadas sobre elementos especificados puramente formales. En la medida en que el programa se define en términos de operaciones computacionales sobre elementos definidos de manera puramente formal, lo que el ejemplo sugiere es que esos elementos por sí mismos no tienen conexiones interesantes con la comprensión. No son ciertamente condiciones suficientes, y no ha sido dada la más mínima razón para suponer que sean necesarias o incluso que hagan una contribución significativa a la comprensión. (Searle, 1980/1981, p. 286). [96-97]

Se intenta que este argumento socave la afirmación de que los científicos cognitivos que intentan comprender la cognición analizando el programa usado por la mente sean capaces de explicar la intencionalidad de los estados mentales. Searle ha defendido su interpretación del *Gedankenexperiment* de la habitación china contra un cierto número de posibles objeciones. Puesto que una de esas objeciones es particularmente probable que se les ocurra a los lectores, merece la pena que la consideremos brevemente. La objeción es que, mientras que Searle no sabe chino, él, junto con las reglas para procesar las preguntas para producir las respuestas, sí lo sabe. Searle ha respondido que el que las reglas sean externas a él es algo accidental —él podría perfectamente memorizarlas—. Pero no comprendería todavía chino. Sólo se comportaría como alguien que entendiese chino. La intuición de Searle parece correcta —la mayor parte de la gente no afirmaría que entiende chino o pretendería que sus respuestas fuesen sobre cosas si operasen de esa manera—. Pero quizás esto sucede porque el *Gedankenexperiment* de Searle representa falsamente el género de reglas que se necesitan para entender un lenguaje. El exige una regla separada para cada pregunta y para cada historieta para la que se ha de dar una respuesta. Tal conjunto de reglas no podría, en principio, proporcionar respuestas a la infinita variedad de preguntas e historietas a las que un chino podría responder. Si pudiésemos habérnoslas con un conjunto de reglas que efectivamente pudiesen bastar para llevar a cabo el género de conversación que Searle ha imaginado, está lejos de ser claro que Searle pudiese convencemos de que el sistema no entiende chino. ¡Las reglas podrían codificar justamente lo que se exige para comprender chino![14].

Aunque el caso que presenta Searle se intenta que cuente contra la adecuación de la Teoría Computacional de la Mente, su afirmación de que un sistema formal es insuficiente para dar cuenta de la intencionalidad no va realmente en contra de ella. Ya hemos visto que la Teoría Computacional dejaba la cuestión de la intencionalidad de los símbolos formales totalmente inexplicada. Lo que es más sorprendente es la afirmación de Searle de que las teorías computacionales no desempeñan ningún papel a la hora de explicar la conducta intencional. Las teorías computacionales se intenta que caractericen el género de procesos internos que ocurren dentro de un sistema y que lo capacitan para que se comporte de la manera apropiada. Si los enfoques computacionales son incorrectos, parece que debe reclamarse [97-98]

alguna explicación de lo que capacita a ciertas clases de sistemas para mostrar intencionalidad. La respuesta de Searle a este problema es la afirmación de que, por defecto, tiene que ser la biología de un sistema lo que proporciona el equipo para exhibir intencionalidad:

> No se debe a que yo sea una instancia de un programa de computador el que yo sea capaz de entender castellano y tenga otras formas de intencionalidad (yo soy, supongo, la instanciación de cualquier número de programas de computador), sino que, hasta donde sabemos, eso se debe a que soy una cierta clase de organismo con una cierta estructura biológica (esto es: química y física), y esta estructura es causalmente capaz, bajo ciertas condiciones, de producir percepción, acción, comprensión, aprendizaje y otros fenómenos intencionales. Y parte del objeto del presente argumento es que sólo algo que tiene esos poderes causales podría tener esa intencionalidad. Quizás otros procesos físicos y químicos podrían producir exactamente esos efectos; quizás, por ejemplo, tienen también intencionalidad, pero sus cerebros están hechos de una materia diferente. Esto es una cuestión empírica, más bien parecida a la cuestión de si la fotosíntesis puede ser hecha por algo con una química diferente de la de la clorofila. (1980/1981, p. 299; ver también Searle, 1984.)

La analogía entre intencionalidad y fotosíntesis devalúa de modo efectivo la posición de Searle. No podemos investigar si una sustancia distinta de la clorofila podría producir la fotosíntesis a menos que sepamos qué capacidades causales capacitan a la clorofila para llevarla a cabo. Searle parecería obligado a reconocer que se desarrollará alguna explicación de lo que tienen que poseer las capacidades interactivas del cerebro para exhibir intencionalidad. Si alguien pudiera desarrollar un análisis de los procesos causales incluidos en la producción de la intencionalidad, se proporcionaría una base para la construcción de una teoría parecida a un programa que describiese esos procesos. Searle, por tanto, tiene simplemente que limitarse a aseverar que los fenómenos intencionales son biológicos y no debe intentar explicar cómo la biología produce la intencionalidad de la misma manera que podemos explicar químicamente cómo ocurre la fotosíntesis. La intencionalidad, por tanto, sigue siendo un misterio de acuerdo con el análisis de Searle.

4.6 EL ENFOQUE DE LA POSTURA INTENCIONAL

Dennett (1971/1978) ha adoptado un enfoque de la intencionalidad que es radicalmente distinto de los que hemos examinado hasta ahora. Afirma que cuando caracterizamos un sistema, ya sea natural [98-99] o artificial, en términos de creencias y deseos, adoptamos lo que él ha llamado «la postura intencional». Ésta es la perspectiva desde la que contemplamos a la gente en la vida diaria, y Dennett mantiene que habrá de ser útil algunas veces el contemplar otros sistemas de una manera parecida. Esta perspectiva no es sólo conveniente cuando estamos intentando predecir cómo se podría comportar una persona u otro sistema, sino que puede también ser útil cuando queremos explicar por qué tal sistema se ha comportado como se ha comportado. Sin embargo, para desarrollar la explicación tenemos que cambiar las perspectivas y adoptar aquello a lo que Dennett se ha referido como la «postura del diseño». Desde la

postura del diseño describimos las actividades mecánicas del sistema que lo capacitan para actuar como un sistema intencional. (En el capítulo 7, al discutir el Funcionalismo Homuncular, describo con mayores detalles la estrategia de Dennett para pasar de la postura intencional a la postura del diseño.)

Aunque manteniendo que la postura intencional en la que caracterizamos los sistemas en términos de creencias y deseos nos es a menudo útil, Dennett ha afirmado también que ningún sistema, incluidos nosotros mismos, es *realmente* intencional. El punto de vista de que las entidades que postulamos son ficticias y no existen realmente se denomina por lo común «instrumentalismo». Entonces, aunque Dennett parece ser un instrumentalista por lo que respecta a las atribuciones intencionales de creencias y deseos, él sólo acepta a regañadientes esta etiqueta. Tiene cierta reluctancia puesto que ha mantenido que no podemos desenvolvemos sin la postura intencional ya sea en la práctica o en principio. Desde la postura intencional, ha afirmado él, adquirimos información que no estaría disponible de otra manera. Además, esta información es «sobre algo perfectamente objetivo: los *modelos* de la conducta humana que se describen a partir de la postura intencional, y solamente a partir de esa postura, y que apoya generalizaciones y predicciones» (Dennett, 1981c, p. 64).

Un aspecto de la discusión por parte de Dennett de la postura intencional la hace aparecer casi vacía. Dennett ha dicho que podemos adoptar la postura intencional hacia casi todo. Por ejemplo, podemos atribuir a una estantería para libros el deseo de mantener los libros en un lugar adecuado y la creencia de que el sostenerlos tal como lo hace cumplirá esa tarea. Este uso de la postura intencional no proporciona información útil. Pero Dennett ha mantenido que, cuando tratamos con sistemas como los de los seres humanos, las atribuciones de creencia y deseo no son tan triviales y la postura intencional proporciona información teórica importante. Nos dice cómo se relaciona el sistema con su entorno: qué información ha adquirido [99-100] y qué acciones está dispuesto a realizar. Esto nos lleva a decir «que el organismo refleja continuamente el entorno, o que hay una *representación* del entorno en —o implícita en— la organización del sistema» (Dennett, 1981c, p, 70). Para que un sistema esté en una relación tal con su entorno, tiene que tener suficientes recursos internos y, por tanto, «el criterio aparentemente frívolo e instrumentalista de la creencia coloca una severa restricción sobre la constitución interna de un creyente genuino, y de este modo proporciona, después de todo, una versión robusta de la creencia» (p. 68).

Dado que él considera que las adscripciones intencionales son útiles, parecería que Dennett debería tratar a las creencias y deseos como reales. (Ver Richardson, 1980, para las razones por las que Dennett debería ser realista por lo que respecta a la postura intencional.) Dennett, sin embargo, ha citado un buen número de razones para no ser realista. Una de las más importantes se inspira en el argumento de Putnam de la Tierra Gemela, que hemos discutido previamente. Para Dennett esto muestra que las adscripciones intencionales son relativas a un entorno y, de este modo, no son caracterizaciones intrínsecas de un sistema. Esto sugiere que aquello a lo que efectivamente se opone Dennett no es a la realidad de los estados intencionales como las creencias y deseos, sino al punto de vista de que éstos son estados internos del sistema. De hecho, Dennett ha dicho que «la creencia es un fenómeno perfectamente objetivo» (Dennett, 1981c, p. 55). Son las teorías computacionales como las de Fodor las que tratan los estados intencionales como estados internos, y así aquello a lo que parece estar oponiéndose Dennett al rechazar el realismo respecto de los estados intencionales es el punto de vista computacional. Esto se pone claramente de manifiesto en un pasaje de la recensión de Dennett (1977) de *The Language of Thought* de Fodor:

> En una conversación reciente con el diseñador de un programa para jugar al ajedrez oí la siguiente critica de un programa rival: «Piensa que debería sacar su reina pronto.» Esto adscribe una actitud proposicional al programa de una manera muy útil y predictiva, pues, como el diseñador continuó diciendo, se puede usualmente contar con comer la reina en el tablero. Pero en ninguno de los muchos niveles de representación explícita que se encuentran en el programa hay algo aproximadamente sinónimo de «Debería sacar mi reina pronto» señalado de manera explícita. El nivel de análisis al que pertenece la observación del diseñador describe rasgos del programa que son, de una manera completamente inocente, propiedades emergentes del proceso computacional que tiene «realidad ingenieril». No veo razón alguna para creer que la relación entre habla de creencia y habla de procesos psicológicos haya de ser más directa (p. 279).

¿Qué más podrían ser los estados intencionales si no son estados [100-101] internos de un sistema? Como he argumentado en otra parte (Bechtel, 1985a), el argumento de Dennett de que las atribuciones intencionales dependen del entorno del sistema sugiere una respuesta. Podríamos interpretar las creencias y otros estados intencionales como estados relacionales que se dan entre un sistema y su entorno. Las atribuciones de creencias y deseos no describirían entonces estados internos de un sistema, sino que describirían cómo se relaciona con un entorno. Un sistema tendría una creencia sobre el agua si estuviese en la relación apropiada con el agua.

Esta propuesta, sin embargo, tiene que puntualizarse. Al discutir la posición de Brentano (capítulo 4 de este volumen) observamos que un punto de vista relacional sobre la intencionalidad resulta problemático puesto que uno de los rasgos importantes de los estados intencionales es que pueden representar fenómenos no existentes. Un sistema no podría posiblemente estar en una relación con algo que no existe. Aunque esto parecería condenar el enfoque que he sugerido, sin embargo no lo hace. Para evitar este obstáculo tenemos primero que adoptar una interpretación holista, no atomista de los estados mentales (en el espíritu de Quine y Davidson; ver capítulo 2). Solamente el conjunto total de los estados cognitivos de una persona es lo que intentamos poner en relación con el entorno. Además, podemos apelar a un concepto de un mundo nocional que Dennett (1982) ha introducido. Dennett introduce esta noción para especificar lo que se representa en el estado mental de una persona. Un mundo nocional no es el mundo efectivo, sino un mundo posible (ver capítulo 2) en el que todas las creencias que una persona tiene serían verdaderas y todos los deseos de ella serían razonables. Para identificar tales mundos Dennett ha propuesto que debemos empezar con el mundo efectivo y considerar cómo podríamos modificarlo para hacer que las creencias falsas de una persona fuesen verdaderas y sus deseos no razonables razonables. Los mundos modificados que cumplen esas condiciones son los mundos nocionales de esa persona.

Los mundos nocionales nos permiten caracterizar relacionalmente los estados intencionales de una persona sin tener que relacionarlos a todos con el mundo actual. Para ver cómo se hace esto, sería útil contemplar los estados mentales de una persona como algo que se puede comparar a los rasgos biológicos. Lo mismo que evaluamos los rasgos biológicos en términos de cuan adaptativo hacen un organismo a un entorno, así podemos evaluar las creencias en términos de cuan adaptado hacen el sistema a su entorno. Lo mismo que algunos rasgos biológicos están bien ajustados al entorno del organismo, del mismo modo algunas creencias serán apropiadas para el entorno del sistema, puesto que los objetos

existen efectivamente de la manera [101-102] especificada. En este caso el enfoque relacional puede aplicarse sin dificultad. Algunos rasgos biológicos no están bien adaptados y, con todo, podemos determinar a qué género de entorno lo estarían. Hacemos algo comparable por lo que respecta a las creencias falsas cuando postulamos un mundo nocional. Aunque en el mundo no hay estados con los que se relacionen esas creencias falsas, podemos decir con qué géneros de estados posibles se relacionarían y cómo difieren de los estados que no existen[15].

El instrumento de los mundos nocionales nos proporciona entonces una manera de interpretar la postulación de Dennett de una postura intencional de un modo realista, no instrumentalista. Las creencias y los deseos caracterizan generalmente a las personas en términos de cómo ellas se relacionan con rasgos de su entorno, y damos cuenta de las diferencias observando cómo sus mundos nocionales se diferencian del mundo efectivo. (He desarrollado este análisis adicionalmente en Bechtel, 1985a.) En este enfoque de Dennett he comparado las propiedades intencionales con propiedades biológicas adaptativas de organismos. Esto sugiere que podríamos incorporar un análisis de la intencionalidad dentro de una armazón evolucionista generalizada[16]. Aunque esto se aparta claramente del instrumentalismo de Dennett, encaja bastante bien con otras características [102-103] de su sistema. Dennett (1978b), por ejemplo, rechaza las constricciones conductistas de B. F. Skinner en contra de postular operaciones mentales inteligentes dentro de la mente, afirmando que postular actividades inteligentes en la mente es aceptable en la medida en que puede darse una explicación evolucionista de cómo la mente puede llegar a adquirir esos procesos inteligentes. También argumenta que la clásica ley evolucionista del efecto (que la conducta puede modificarse de acuerdo con si es premiada o castigada) es simplemente una forma internalizada de selección natural que es ella misma el producto de la selección natural[17] (Dennett, 1975/1978).

El tratar las adscripciones intencionales de creencias y deseos a un sistema como una caracterización de la relación entre el sistema cognitivo y su entorno tiene algunas consecuencias importantes para la ciencia cognitiva. Argumenta a favor de a) diferenciar nuestras caracterizaciones intencionales de los sistemas cognitivos de modelos de procesamiento interno, pero también b) a favor de comprender la cognición en su contexto de entorno y filogenético. Aquí desarrollo brevemente estos puntos.

La primera consecuencia es algo que ya hicimos notar al distinguir la Teoría Representacional de la Mente de la Teoría Computacional. Ahora podemos ver más claramente por qué las adscripciones intencionales de creencias y deseos deberían distinguirse de los modelos de procesamiento interno. Las actitudes preposicionales son [103-104] un modo de caracterizar el sistema cognitivo con relación a su entorno, pero no es poco común en ciencia usar explicaciones diferentes para describir el comportamiento de un sistema y para describir los procesos internos que hacen posible la conducta. Por ejemplo, una célula de levadura que está realizando la fermentación se describe fisiológicamente como metabolizando azúcar para producir alcohol, mientras que en bioquímica la reacción se explica en términos de redes de enzimas y factores adicionales que, todos juntos, hacen posible que una célula metabolice azúcar. Similarmente, podemos contemplar la caracterización de cómo un sistema se relaciona con su entorno como algo diferente del modelo de procesamiento que explica cómo es capaz de llevar a cabo esto. Cuando intentamos desarrollar de modo efectivo un modelo de procesamiento, hay diferentes tipos que podríamos considerar, incluyendo el modelo computacional tal como ha sido articulado por Fodor y se ha empleado en IA tradicional, un modelo sintáctico como el descrito por Stich

y que ha sido empleado en muchos de los trabajos tradicionales en psicología del procesamiento de la información, o un modelo conexionista como han defendido algunos teóricos recientes y como se ha investigado en IA recientemente. La adecuación del modelo de procesamiento está determinada por si describe correctamente el proceso que opera en los sistemas cognitivos reales, no por si invoca la estructura formal de las explicaciones que describen la conducta del sistema cognitivo en su entorno.

Aunque podemos entonces distinguir la tarea de desarrollar explicaciones intencionales que invoquen actitudes preposicionales de la de desarrollar explicaciones de procesamiento, esta perspectiva sobre la intencionalidad sugiere también modos en los que las dos armazones necesitan estar relacionadas. Es importante para aquellos que trabajan en explicaciones sobre el procesamiento el prestar atención a la perspectiva intencional, en la que la conducta de un sistema cognitivo se caracteriza en términos de sus creencias y deseos sobre el entorno. Esta perspectiva intencional es la que identifica aquellos aspectos de la conducta de un sistema que necesitan ser explicados por el enfoque del procesamiento. (Lo que se requiere es a lo que Darden y Maull, 1977, se refieren como una «teoría intercampos». Ver Bechtel, en prensa b, capítulo 6, para más detalles sobre las teorías intercampos).

Desde esta perspectiva podemos dar sentido a las llamadas de psicólogos como J. J. Gibson (1969) y Ulric Neisser (1975, 1982) para adoptar una perspectiva ecológica en psicología. Ambos objetan el excesivo énfasis en el trabajo de laboratorio en psicología (p. ej., los estudios de memoria con sílabas sin sentido o los estudios de visión [104-105] que usan estímulos presentados taquistoscópicamente) que consideran que no se concentran en los rasgos realmente importantes de los sistemas cognitivos. Tanto Gibson como Neisser argumentan que en sus hábitats naturales los organismos no responden a los estímulos simples usados en investigación de laboratorio, sino a conjuntos coherentes de estímulos que tienen a la vez extensión espacial y duración temporal. Gibson llamó a esos estímulos «facilitadores» («*affordances*») puesto que presentan información que facilita la acción a los organismos.

La perspectiva intencional es similar a la perspectiva ecológica de Gibson y Neisser en la medida en que se concentra en la información del entorno a la que está respondiendo el sistema. Pero, si recordamos el enfoque de Dennett de la relación entre la postura intencional y la postura del diseño, podemos ver también cómo se relacionaría la perspectiva intencional con las explicaciones del procesamiento de información en el nivel del diseño. No necesitamos dar el paso adicional que Gibson dio cuando emparejó su exigencia de un enfoque ecológico con un repudio del enfoque del procesamiento de la información. (Ver Fodor y Pylyshyn, 1981, y Hamilyn, 1977, para argumentaciones a favor de que el procesamiento de información se requiere aun si se aceptan ciertos aspectos de la posición de Gibson.) Hay procesos internos que capacitan a un sistema cognitivo para tener estados intencionales, y se necesita trabajo de laboratorio para identificarlos. Lo que hace la postura intencional es proporcionar una perspectiva para identificar cómo se relaciona el sistema con su entorno. Partiendo de esta perspectiva, esa investigación de laboratorio puede identificar qué procesos internos lo capacitan para que se relacione de la manera que lo hace. (Ver Glotzbach y Heft, 1982, para un argumento relacionado.)

Dennett (1983) propone que la postura intencional ofrece una armazón para la etología cognitiva, una disciplina que busca identificar las capacidades cognitivas de organismos particulares (y por extensión, quizás, de sistemas artificiales) que son relevantes en sus hábitats naturales. La etología cognitiva puede generar aquello a lo que Anderson (1986) se ha referido como «perfil cognitivo» de una especie. Este perfil proporciona una descripción de los diferentes géneros de información a los que es sensible un

organismo, los géneros de cosas que puede recordar, y los modos en los que puede usar esta información. Por tanto, ofrece una perspectiva del organismo que está entre las explicaciones específicas de cómo se comporta un organismo en el entorno y el procesamiento interno que produce la conducta. La información recogida en el perfil cognitivo dice entonces al investigador que intenta desarrollar los modelos de procesamiento [105-106] interno qué capacidades necesitan explicarse al dar cuenta del procesamiento.

Adoptar el punto de vista de que las adscripciones caracterizan organismos en términos de sus creencias y deseos sobre sus entornos nos permite también colocar nuestros análisis de sistemas particulares en una perspectiva filogenética. Podemos examinar diferentes modos en los que han evolucionado los organismos para relacionarse con su entorno. En el caso de los humanos el lenguaje desempeña claramente un papel de gran importancia a la hora de codificar nuestras creencias sobre nuestro entorno y de representar nuestros deseos. Esto plantea la cuestión de hasta qué punto la intencionalidad de los estados mentales depende de la disponibilidad del lenguaje como un vehículo para la comunicación. Los filósofos han ofrecido una gran variedad de perspectivas sobre la cuestión de si el lenguaje es un prerrequisito de la intencionalidad o hace uso de intencionalidad anterior (p. ej., ver Bennett, 1976; Chisholm, 1984; Gauker, 1987; McDowell, 1980; Sellars, 1963a; Tennant, 1984); los psicólogos han proporcionado también algunas veces evidencia relevante (p. ej.,Furth, 1966).

El interés en si la intencionalidad de los estados mentales es más básica que la del lenguaje ha sido estimulado por el trabajo reciente sobre la comunicación animal, especialmente la investigación sobre el lenguaje llevada a cabo con monos. Aunque ciertamente este trabajo ha sido controvertido, las investigaciones llevadas a cabo por los Gardner (Gardner y Gardner, 1969) y otros sugerían que los chimpancés podrían usar elementos lingüísticos intencionalmente. Este hallazgo podría interpretarse como evidencia a favor de la afirmación de que la capacidad de intencionalidad existe de manera anterior al aprendizaje del lenguaje. Sin embargo, una objeción muy común a los primeros proyectos sobre el lenguaje de los monos era que se requería una configuración de la conducta muy intensa antes de que los animales pudieran usar los símbolos lingüísticos, y no estaba claro que los chimpancés estuvieran usando realmente esos símbolos con significado. Esto arruinaría la afirmación de que los animales ya poseían intencionalidad. Savage-Rumbaugh (1986) proporciona, sin embargo, evidencia completamente compulsiva de que los chimpancés están usando los símbolos intencionalmente. Además, ella está ahora embarcada en un proyecto de investigación pionero con chimpancés enanos (*Pan paniscus*), que demuestra que los miembros de esta rara especie, cuando se les proporciona un entorno adecuado, son capaces de adquirir el uso de símbolos con significados específicos sin un régimen de reforzamiento específico e incluso observando simplemente el uso por los humanos (Savage-Rumbaugh, McDonald, Sevick, Hopkins y Rupert, 1986). [106-107]

La cuestión de si esto indica intencionalidad anterior es, sin embargo, compleja puesto que los chimpancés enanos exhiben también un conjunto razonablemente extenso de vocalización en sus hábitats nativos. Esas vocalizaciones pueden ser ya modos intencionales de vocalización y proporcionar las bases para la capacidad del animal de usar lenguajes más complejos en situaciones experimentales. Por otra parte, otros investigadores, como Carolyn Ristau (1983, 1987), han intentado demostrar que la conducta intencional que se encuentra en animales, como las aves marinas, es claramente no lingüística. Aunque existen cuestiones fundamentales que deben plantearse sobre cómo valoramos la intencionalidad de tales animales, esta investigación sugiere que podemos ser capaces de examinar el desarrollo de la

intencionalidad filogenéticamente mirando cómo diferentes organismos han desarrollado diferentes capacidades para habérselas con la información de su entorno. Un beneficio de tal perspectiva comparativa es que comprender los géneros de capacidades cognitivas a partir de las que se desarrollan nuestras capacidades puede ayudamos a caracterizar adecuadamente nuestras capacidades cognitivas y proporcionamos una guía cuando intentamos explicar qué procesamiento interno hace posible esas capacidades cognitivas.

4.7 RESUMEN DE LOS ENFOQUES FILOSÓFICOS DE LA INTENCIONALIDAD

Los últimos dos capítulos se han concentrado en lo que muchos consideran que es el rasgo definitorio de los estados mentales: su intencionalidad. En el capítulo anterior presenté varios intentos diferentes por parte de los filósofos de decir lo que es distintivo de la intencionalidad y por qué la intencionalidad convertiría en imposible la explicación científica de los estados mentales. En este capítulo he presentado diversas propuestas que los filósofos han avanzado para explicar la intencionalidad dentro de la armazón de la ciencia natural. Comencé con la Teoría Computacional de la Mente, que pretende usar el formato de la actitud proposicional para describir los estados mentales como la base para generar una explicación del procesamiento interno. Así, Fodor propone una teoría de las actividades psicológicas que postula que la gente realiza efectivamente inferencias en un lenguaje del pensamiento. Este enfoque es común en IA, pero no explica la intencionalidad. Presenté a continuación la Teoría Representacional de la Mente como una posición que mantenía que la mente era intencional y se describía apropiadamente en términos de actitudes preposicionales pero rechazaba la idea de que [107-108] el procesamiento interno incluyese el procesamiento de esas proposiciones.

He discutido tres modos en que los filósofos han intentado explicar las capacidades representacionales de la mente: el enfoque de la teoría de la información de Dretske, la reducción biológica de Searle, y el enfoque de la postura intencional de Dennett. El enfoque de Dretske usaba la teoría matemática de la información para explicar cómo un estado podría ser sobre otro. Tenía la virtud de convertir la intencionalidad en un fenómeno natural, pero parecía problemática en tanto que trataba principalmente a las capacidades cognitivas como si introdujesen distorsiones dentro del proceso, de otra manera verídico, de adquirir conocimiento. Una perspectiva evolucionista sugeriría que los estados mentales desempeñan un papel más positivo al generar la intencionalidad. El enfoque de Searle ligaba la intencionalidad a nuestra constitución biológica, pero parecía hacer de ella algo misterioso. Afirmaba que la intencionalidad era un fenómeno biológico, pero negaba que pudiésemos explicar lo que hace que ciertos estados biológicos sean intencionales. La perspectiva de Dennett de la postura intencional hacía de la perspectiva intencional algo que adoptamos con respecto a ciertos sistemas. Lo que parecía más problemático en este enfoque era su instrumentalismo con respecto a las atribuciones intencionales, pero he sugerido cómo podríamos desarrollar una versión del enfoque de Dennett que contemplase los estados intencionales de una manera realista. Hace esto al tratarlos como estados del sistema que se adaptan a rasgos del entorno del sistema.

Brentano pensaba que la intencionalidad de los estados mentales tenía implicaciones para el género de entidad que consideramos que son las mentes. Las mentes, afirmaba él, no podrían ser cuerpos físicos puesto que los objetos físicos carecían de intencionalidad. Muchos de los filósofos cuyos puntos de vista

hemos discutido en este capítulo han intentado, sin embargo, mostrar cómo los estados intencionales podrían surgir en sistemas físicos. Pero esto apunta a una cuestión fundamental: ¿Cuál es la relación entre las mentes y los objetos físicos? Éste es el punto central de los dos capítulos que siguen. [108-109]

5. EL PROBLEMA MENTE-CUERPO: DUALISMO Y CONDUCTISMO FILOSÓFICO

5.1 INTRODUCCIÓN

Durante tres siglos la investigación filosófica se ha centrado en dos preguntas sobre las mentes: qué géneros de cosas son las mentes y cómo se relacionan las mentes con los cuerpos. En este capítulo y en el capítulo 6 me propongo explorar las posiciones principales que los filósofos han avanzado para responder a esas preguntas. Mi discusión seguirá generalmente el orden en que se desarrollaron esas posiciones puesto que las últimas posiciones se propusieron generalmente para vencer las dificultades a las que, se pensaba, tenían que hacer frente las primeras. No debe concluirse de esto que las posiciones discutidas en primer lugar tienen sólo interés histórico puesto que, a pesar de todo, cada una de esas posiciones dispone todavía de activos defensores tanto entre los filósofos como entre los practicantes de las ciencias cognitivas. Comienzo este capítulo con una discusión del dualismo mente-cuerpo, que ha servido como el punto de partida principal para aquellos que han desarrollado posiciones alternativas. Examino también el conductismo filosófico, que constituye uno de los primeros intentos de evitar el dualismo e integrar los fenómenos mentales dentro del universo físico.

5.2 DUALISMO

El término *dualismo* se aplica generalmente a posiciones que contemplan los fenómenos mentales como fuera de alguna manera de la armazón de la ciencia natural. Necesitamos distinguir dos amplias clases de dualismo: dualismo de substancias y dualismo de propiedades. El *dualismo de substancias* considera que la mente es una entidad no física separada del cuerpo. El *dualismo de propiedades* es una posición más modesta que no postula entidades no físicas pero que mantiene que algunas de las propiedades que poseen esos objetos constituyen una clase distinta de propiedades mentales. El dualismo de substancias es la posición mejor conocida y será la forma principal que discutiré en esta sección. [109-110]

La cuestión misma de si la mente es una substancia diferente del cuerpo físico es un legado de Descartes. Ahora bien, el cuadro cartesiano está tan arraigado en nuestra cultura general que muchas personas encuentran difícil concebir una alternativa donde la cuestión no surja. Sin embargo, la diferenciación entre la mente y el cuerpo era completamente extraña a la perspectiva aristotélica que precedió a Descartes. El enfoque aristotélico caracterizaba y clasificaba objetos en términos de lo que hacían más bien que en términos de su carácter intrínseco. Esto es quizás una diferencia sutil, pero lleva a formas radicalmente diferentes de investigación. Como vimos en el capítulo 1, Aristóteles distinguió entre la *materia* y la *Forma* de un objeto, pero mantuvo que cualquier objeto consistía en la materia organizada de acuerdo con una Forma particular. Aristóteles concentraba su atención en la Forma, no en la materia, puesto que un objeto se caracterizaba en virtud de la Forma. Esto se aplicaba no solamente a

objetos inanimados, sino también a objetos animados. Aristóteles habló de la Forma de las cosas vivientes como su *psique* o *alma*, pero no pensaba en ella como una parte discreta del organismo viviente. Más bien, él la contemplaba como carácter definidor del organismo.

Para Aristóteles la Forma, tanto de los objetos animados como inanimados, se descubre observando el género de actividades que realizan. Aristóteles distinguió tres clases de organismos en términos de las actividades que son capaces de realizar y, por consiguiente, identificó tres géneros diferentes de almas. Las plantas son capaces de tomar nutrientes y reproducirse, y esas funciones definen el alma vegetativa. Los animales no sólo son capaces de esas actividades, sino que son capaces también de sentir las cosas de su entorno y de moverse en él, y esas funciones definen el alma animal. Finalmente, los humanos son también capaces de razonar, que es la función distintiva de sus almas (ver *De Anima* en McKeon, 1941).

Dentro del pensamiento aristotélico no hay virtualmente tentación de pensar en el alma como una cosa distinta que podría estar separada del resto del organismo. (La puntualización «virtualmente» tiene que añadirse, puesto que parece que Aristóteles, al menos, juega con la idea de que el alma pensante podría ser capaz de sobrevivir a la disolución del cuerpo.) La revolución científica de los siglos XVI y XVII tuvo como resultado el rechazo de la explicación aristotélica de la naturaleza en términos de materia y Forma, y esto llevó últimamente a una perspectiva diferente sobre la actividad mental. La nueva física tenía como concepción básica que la materia era pasiva e inerte, sujeta a las fuerzas que incidían sobre ella desde fuera. La tarea de la física era desarrollar las leyes que gobiernan las maneras [110-111] en que los objetos se afectan entre sí, bien golpeándose o ejerciendo fuerzas sobre ellos. Surgió la cuestión de si este punto de vista podía extenderse también a las actividades de los animales y de los humanos. Muchos investigadores pensaron que debía hacerse. El filósofo del siglo XVII Thomas Hobbes es quizás el mejor conocido de los que presionaron a favor de una explicación completa de la actividad humana, incluyendo el pensar, en los mismos términos en que se hacía para los objetos físicos no animados. Incluso Descartes estaba fuertemente atraído por esta expectativa. El estaba fascinado por la conducta de los sistemas hidráulicos y los contempló como posibles modelos de los procesos fisiológicos en los humanos y en otros animales. El trabajo de Harvey sobre la circulación de la sangre, que incluía una bomba que empujaba el fluido a lo largo de una serie de canales, era un modelo fácilmente disponible para Descartes. Descartes defendió un punto de vista similar para el sistema nervioso, interpretándolo como un sistema de canales a través de los cuales los humores animales estaban circulando. Esta circulación, pensaba él, producía mecánicamente la conducta física de los sistemas vivientes.

Descartes, sin embargo, mantenía que este intento de explicar la conducta en términos físicos alcanzaba un límite inevitable en aquellos asuntos humanos que incluían el uso del lenguaje y del razonamiento. Consideraba que esas actividades humanas eran tan diferentes en género de las que se encontraban en el resto de la naturaleza, que no pensaba que pudiesen explicarse de la misma manera. Él no negaba que los sistemas mecánicos u otros animales (que él consideraba que eran simplemente sistemas mecánicos) podían emitir palabras, pero afirmaba que «jamás sucede que [un animal no humano] dispone sus palabras de varias maneras para responder apropiadamente a todo lo que puede decirse en su presencia, como puede hacer incluso el tipo menos instruido de hombre» (Descartes, 1637/ 1970, p. 116). Respecto al razonamiento, pensaba que, aunque las máquinas o los animales podrían conducirse apropiadamente en muchos contextos específicos, no exhibirían el tipo de racionalidad general que los humanos exhiben. Esas diferencias entre los humanos y otros animales, pensaba

Descartes, podrían explicarse solamente si postulamos un género especial de substancia en los seres humanos: la substancia mental.

Una substancia se caracteriza para Descartes por una propiedad básica de la que no puede carecer y seguir siendo con todo la misma substancia. Para la substancia física esta propiedad es la extensión (esto es: la ocupación de espacio). Descartes afirmaba que, aunque podemos imaginar que otras características de los objetos físicos cambian o se eliminan radicalmente, tenemos siempre que interpretarlos [111-112] como ocupando alguna cantidad de espacio. En contraste con la substancia física, Descartes consideró que la propiedad definitoria de la substancia mental era el pensar. Descartes interpretó el pensar genéricamente de modo que incluyese creer, suponer, esperar, y así sucesivamente. (Descartes incluye aquí la misma clase de actividades que describiríamos en el discurso de actitudes preposicionales y que Brentano describiría como intencionales. Ver capítulo 3.) Descartes mantuvo que pensar y extensión definen dos clases diferentes de objetos. La naturaleza radical del desdoblamiento que Descartes diseñó se torna claro en sus *Meditaciones acerca de la Filosofía Primera*. Después de arrojar dudas sobre tantas de sus creencias como le resultaba posible, Descartes concluyó inicialmente que sólo su creencia de que él existe como una cosa pensante está más allá de toda duda. Aunque era capaz de dudar que tenía un cuerpo, no era capaz de dudar de que tenía una mente. Puesto que Descartes podía imaginar su mente existiendo sin su cuerpo, concluía que hay dos géneros totalmente separados de entidades.

Contra el dualismo de Descartes se han lanzado múltiples objeciones. Una de las más serias se centra en la interacción entre la mente y el cuerpo. Si las dos substancias son tan diferentes, parece difícil explicar cómo pueden interactuar entre ellas: ¿cómo podrían los pensamientos causar movimientos físicos del cuerpo? Descartes propuso una solución. Afirmaba que en un lugar central del cerebro —la glándula pineal— la mente podía alterar los movimientos de los humores animales fluyendo a través de los canales nerviosos e influyendo de esta manera en la actividad del cuerpo. Aunque la investigación subsiguiente ha desacreditado a la teoría de los humores animales y ha identificado una función distinta para la glándula pineal, éstos no son los problemas más serios para la solución propuesta por Descartes. Queda el problema más básico de explicar cómo dos substancias que difieren tan radicalmente pueden afectarse mutuamente. Gassendi planteó la objeción como sigue:

> Queda aún por explicar cómo esa unión y aparente entremezcla [de la mente y el cuerpo]... puede encontrarse en ti, sí eres incorpóreo, inextenso e indivisible... ¿Cómo, al menos, puedes unirte con el cerebro, o con alguna diminuta parte de él que (como se ha dicho) tiene que tener con todo alguna magnitud o extensión, por pequeña que sea? Si tú careces completamente de partes, ¿cómo puedes mezclarte o parecer mezclarte con sus diminutas subdivisiones? Pues no hay mezcla alguna a menos que cada una de las cosas que han de mezclarse tenga partes que puedan mezclarse entre sí. (Gassendi. 1641/1970. p. 201.)

La misma cuestión fue planteada por Descartes a la princesa Isabel en 1643: «¿Cómo puede el alma del hombre, que es sólo una [111-112] substancia pensante, determinar que sus humores corporales realicen acciones voluntarias?» (Kenny, 1970, p. 135).

Descartes mantuvo que tales objeciones eran ilegítimas. En primer lugar, suponían que la interacción de la mente y del cuerpo debía seguir el modelo común de interacción causal cuando realmente incluye

una clase completamente diferente de interacción. En segundo lugar, él defendía que «la mente humana [no] es capaz de concebir al mismo tiempo la distinción y la unión entre mente y cuerpo, puesto que para ello es necesario concebirlas como una cosa simple y al mismo tiempo concebirlas como dos cosas; y esto es absurdo» (Kenny, 1970, p. 142). Muchos de los comentadores encuentran insuficientes las respuestas de Descartes. Sin embargo, Richardson (1982) ha mantenido que, lógicamente hablando, son suficientes. Él afirma que en su primera respuesta Descartes estaba observando que cualquier explicación en términos de fuerzas debe, en última instancia, detenerse en algunas fuerzas que han de considerarse como básicas, y así defendía que tenemos que detener la búsqueda de la explicación de la interacción postulando la existencia de un modo de interacción causal entre la mente y el cuerpo. Para explicar la segunda respuesta Richardson ha apelado a la repetida negativa por parte de Descartes a considerar la relación entre mente y cuerpo como comparable a la de un piloto y un barco. Más bien, él consideró esta relación como algo mucho más íntimo. Richardson ha propuesto que Descartes trata algunos estados como estados de las substancias unidas (así pues, entidades con dos naturalezas) y no de ninguna de las dos sola. Él cita como evidencia el siguiente pasaje de Descartes:

> hay... ciertas cosas que experimentamos en nosotros mismos y que no deberían ser atribuidas ni a la mente ni al cuerpo solos, sino a la estrecha e íntima unión que existe entre el cuerpo y la mente... Tales son los apetitos de hambre, sed, etc., y también las emociones o pasiones de la mente que no subsisten en la mente o en el pensamiento solos... y finalmente todas las sensaciones. (Descartes, 1644/1970, p. 238.)

Si esos estados lo son de una substancia unida, entonces, en la medida en que son partes de una substancia física, pueden interactuar con las substancias físicas de la manera ordinaria. De igual manera, en tanto que son estados de una substancia mental, pueden interactuar con otros estados mentales de la manera apropiada a los estados mentales. Aunque esto hace que la respuesta de Descartes parezca más coherente de lo que generalmente se ha pensado que es, queda sumido en un gran misterio el explicar cómo las dos naturalezas pueden combinarse para formar una entidad. Así pues, el debate sobre cómo puede ocurrir la interacción entre mente y cuerpo continúa. [113-114]

Aunque Descartes se contempla a menudo como el dualista paradigmático, ha habido muchos otros desde su época. Brentano y William James son dos prominentes dualistas del siglo XIX. En nuestros propios días el filósofo Karl Popper y el neurofisiólogo John Eccles han avanzado conjuntamente una versión del dualismo (en realidad un tri-ísmo) que prefieren denominar «interaccionismo» (Popper y Eccles, 1977). Al igual que Descartes, se concentran en aspectos de la actividad mental que, afirman ellos, no pueden ser llevados a cabo por cuerpos físicos. Uno de tales aspectos es la capacidad de las actividades mentales para generar objetos abstractos de pensamiento que asumen una vida por sí mismos. Aquí se incluyen los objetos matemáticos, las teorías científicas y las obras literarias. Popper[1] caracterizó esos objetos como constituyentes de un mundo separado que él llamó el «Mundo 3». El Mundo 3 se distingue del Mundo 1 —el mundo de los objetos físicos— y del Mundo 2 —el mundo de la actividad mental— por el hecho de que está gobernado por principios normativos tales como las reglas de la lógica. Popper ha insistido en que los principios de la lógica tienen una validez objetiva, sígalos o no alguien alguna vez, y de este modo postula que tienen una existencia objetiva en un mundo separado

del mundo físico o del mundo del pensamiento.

El argumento de que las actividades mentales son distintas de las actividades físicas se sigue de la necesidad de un intermediario que pueda aplicar información del Mundo 3 al Mundo 1 físico. Popper ha afirmado que ningún sistema puramente físico puede captar los contenidos abstractos del Mundo 3. Por tanto, tiene que haber actividades mentales que capten objetos del Mundo 3 y, a continuación, interactúen causalmente con eventos del Mundo 1. Los críticos se han opuesto a la afirmación de Popper de que ningún objeto del Mundo 1 puede interactuar con objetos abstractos. Los no dualistas mantienen que no hay nada problemático en que los objetos físicos capten objetos abstractos. Incluso sistemas claramente físicos como los computadores pueden diseñarse de tal manera que sigan las reglas de la lógica y razonen sobre teorías científicas u obras literarias. La clave para que sean capaces de hacer esto es su diseño, pero este diseño se encuentra en su existencia física y no es algo distinto (ver P. S. Churchland, 1986, p. 340)[2]. [114-115]

Para apuntalar su posición. Popper ha desarrollado un argumento que intenta mostrar que sólo el interaccionismo puede dar una explicación *apropiada* de cómo el Mundo 3 regula las actividades del Mundo 1. El término *apropiado* resulta crítico en este contexto, puesto que Popper admite que los objetos del Mundo 3 son frecuentemente instanciados en objetos del Mundo 1 (p. ej., una novela resulta instanciada en el papel y en la tinta de un libro) y. por tanto, pueden afectar a otros objetos del Mundo 1 de la manera en que los objetos del Mundo 1 normalmente afectan a los otros objetos del Mundo 1 (p. ej.. sujetando papeles en los que está localizada, etc.). El modo de interacción por el que Popper se preocupa incluye objetos del Mundo 3 que afectan a objetos del Mundo 1 no a causa de su instanciación, sino a causa de sus contenidos. Este argumento se presenta como parte de las teorías fisicalistas discutidas en el capítulo siguiente. Él mantiene que tales teorías han de negar que hay eventos mentales o convertirlos en algo ineficaz:

> Podemos dividir a aquellos que mantienen la doctrina de que los hombres son máquinas, o una doctrina similar, en dos categorías: aquellos que niegan la existencia de eventos mentales, o experiencias personales, o de conciencia; ... y aquellos que admiten la existencia de eventos mentales, pero aseveran que son «epifenómenos», que todo puede explicarse sin ellos, puesto que el mundo material está cerrado causalmente. (Popper y Eccles, 1977, p. 5.)

Puesto que es implausible negar de modo absoluto la ocurrencia de eventos mentales, la única posición plausible para un fisicalista es, de acuerdo con Popper, el epifenomenalismo. El *Epifenomenalismo* mantiene que los estados mentales están apareados con estados cerebrales, pero mantiene que no hay relaciones causales entre ellos. Sólo los estados cerebrales tienen eficacia causal y, de este modo los estados mentales son mudos.

Una vez que se ha interpretado el fisicalismo como una clase de [115-116] epifenomenalismo, Popper afirma que el epifenomenalismo es inconsecuente con la teoría evolucionista, puesto que, de acuerdo con él, la teoría evolucionista está obligada a explicar todos los rasgos de la especie en términos de selección natural. Pero la selección natural puede sólo explicar la emergencia de un rasgo mostrando cómo su posesión proporciona a los individuos de la especie instrumentos para la supervivencia (Popper y Eccles, 1977, p. 73). Puesto que el epifenomenalismo convierte en ineficaz a la actividad mental y, de

este modo, en un instrumento inútil para la supervivencia, la teoría evolucionista no puede explicar el origen de la actividad mental. Dado que Popper considera que la teoría evolucionista es la única que da una explicación plausible de cómo podrían emerger los rasgos, su conclusión es que la posición fisicalista es insostenible.

Este argumento es seriamente defectuoso. Como discuto en el capítulo siguiente, muchos fisicalistas, especialmente los proponentes de la Teoría de la Identidad, rechazarían el tratamiento de Popper de su posición en tanto que entraña epifenomenalismo. Ellos mantienen que los estados mentales son simplemente estados físicos y como tales proporcionan los mismos beneficios que los estados físicos proporcionan (Mortensen, 1978). Pero incluso si concedemos la interpretación de Popper del fisicalismo, su argumento no da en el blanco. Las teorías evolucionistas han propuesto mecanismos distintos de los de la selección natural para explicar el cambio evolucionista (Gould y Lewontin, 1979). Además, incluso si nos restringimos a la selección natural, el argumento falla. La selección natural permite que un rasgo que está ligado a rasgos ventajosos resulte favorecido incluso si él mismo no es ventajoso. Un caso biológico simple ilustra este punto. Explicamos por qué las plantas son verdes no mostrando ninguna ventaja que se siga de su ser verdes, sino mostrando que el alelo responsable de la clorofila en las plantas es también responsable de su color verde y mostrando que poseer clorofila es ventajoso. No exigimos una teoría evolucionista para explicar ni por qué las plantas son verdes ni por qué tienen clorofila, ni siquiera por qué la clorofila causa que las plantas sean verdes. Volvemos a la bioquímica para explicar esta conexión: todo lo que se requiere que haga la teoría evolucionista es explicar por qué el tener clorofila beneficia a las plantas (ver Bechtel y Richardson, 1983). Así pues, incluso si los estados mentales son epifenoménicos respecto de ciertos géneros de estados cerebrales, podrían resultar favorecidos por la selección si esos estados cerebrales ayudasen a los organismos en su búsqueda por la supervivencia. Por consiguiente, el fisicalismo no es inconsecuente con la teoría evolucionista y no estamos forzados a adoptar el interaccionismo como única alternativa. [116-117]

Los argumentos de Descartes y de Popper son los argumentos más comunes a favor del dualismo, pero se han avanzado algunos otros (p. ej., Polten, 1973). Muchas personas son llevadas al dualismo preguntando: ¿Cómo podrían los rasgos de la mente que observamos en la introspección ser explicados en términos de procesos físicos? Por introspección observamos el carácter cualitativo de nuestra vida mental, que parece estar rellena de imágenes, sentimientos y cosas por el estilo. Parece caracterizarse también por la intencionalidad intrínseca (ver discusión de Searle en el capítulo anterior). Esas características parecen ajenas al universo físico de modo que hay una inconmensurabilidad entre lo que reconocemos en nosotros mismos cuando percibimos un objeto y las actividades neurales que ocurren en nuestro cerebro. Igualmente, parece haber una inconmensurabilidad entre el que otra persona haga referencia a un perro y el modelo de actividad neural en el cerebro de esa persona.

Los no dualistas responden comúnmente a esas afirmaciones señalando otras inconmensurabilidades que existen en la naturaleza, tales como las que se dan entre los fenómenos vivientes y los no vivientes. Afirman que, aunque una vez pareció inconcebible el que la materia inerte pudiese manifestar las características de la vida, este vacío ha sido puenteado por la biología moderna. Además, la introspección puede no decirnos de manera fiable cómo son las cosas. Lo mismo que sabemos que nuestros mecanismos perceptivos no revelan la naturaleza esencial del mundo externo, es posible que la introspección no revele la naturaleza real de la experiencia interna. El progreso en la construcción de

máquinas que simulen la conducta humana puede llevarnos también a comprender lo que está realmente implicado cuando hacemos introspección sobre nuestra experiencia.

Merece la pena señalarse en este punto que algunas personas extraen su apoyo al dualismo de una esfera completame diferente. Ven la perspectiva dualista como algo esencial para la comprensión del *status* moral y religioso de los seres humanos. Para mucha gente, nuestra perspectiva moral exige que los agentes humanos sean libres puesto que los juicios morales sólo tienen sentido si los agentes son libres de elegir acciones de acuerdo con sus propias voliciones. En la medida en que cualquier forma de fisicalismo parecería ser determinista al colocar a los seres humanos bajo el control de las fuerzas causales de la naturaleza, el fisicalismo parece eliminar la potencialidad de la libertad humana y, de este modo, nuestra perspectiva moral. Nuestro sistema de juicios morales parece exigir, por tanto, el dualismo.

A este argumento se ha respondido con una gran variedad de réplicas. Una respuesta consiste simplemente en rechazar la afirmación [117-118] de que nuestras perspectivas morales dependen de la libertad humana, tal como hace B. F. Skinner (1948,1971). Otra consiste en argumentar que la forma de libertad que es fundamental para nuestra perspectiva moral no es incompatible con el fisicalismo. Es más, hay una concepción filosófica conocida como *determinismo débil* que mantiene que el libre albedrío y el determinismo son compatibles. Esta posición mantiene que la forma de libertad necesaria para la moralidad es libertad suficiente de constricciones externas de modo que seamos capaces de hacer lo que elegimos hacer (esté o no nuestra elección determinada). Cuando se cumple esta condición, podemos mantener que es posible dar moralmente cuenta de nuestras acciones. No es necesario además que el procedimiento por el que llegamos a nuestra elección sea libre. (Para una exploración filosófica reciente de este problema, ver Dennett, 1984b, 1984c.)

Me he centrado en esta sección hasta ahora en el dualismo de substancias, pero, como he observado al principio, existe una forma más débil de dualismo: el dualismo de propiedades. El dualismo de propiedades mantiene que algunos tienen propiedades mentales además de las propiedades físicas. El trazar esta distinción entre propiedades mentales y físicas permite al dualista captar una intuición compartida por muchos dualistas —que existe un carácter distintivo de los fenómenos mentales— y, con todo, rechazar el objeto de la afirmación del dualista de que tenemos que postular una substancia separada para capturar esta diferencia. Los dualistas de propiedades insisten solamente en que las propiedades mentales son diferenciables de las propiedades físicas. Sin embargo, el mismo objeto es capaz de poseer ambos géneros de propiedades.

Hay, en efecto, distintas versiones del dualismo de propiedades que difieren entre sí en su explicación de cómo se relacionen las propiedades físicas con las mentales. Una versión mantiene simplemente que cada instancia de una entidad que instancia una propiedad mental es una instancia de una propiedad que instancia una propiedad física, sin que haya ninguna otra conexión. Este punto de vista está estrechamente relacionado con la Teoría de la Identidad como Instancia (*Token Identity Theory*), que se discute en el capítulo 6. Una versión más clásica del dualismo de propiedades, la teoría del aspecto dual avanzada por Huxley en el siglo XIX, mantiene que algunos eventos tienen dos aspectos. Generalmente, este punto de vista ha abrazado el epifenomenalismo y mantiene que el aspecto mental del evento no tiene efecto sobre el aspecto físico, aunque se mantuvo algunas veces que el aspecto físico causaba el aspecto mental. De acuerdo con este punto de vista, las propiedades mentales tienen la misma relación relativa

con las operaciones de una persona [118-119] que los *displays* de un CRT tienen con las operaciones que suceden en un computador—simplemente relatan lo que está sucediendo sin influir en el curso de los eventos—. Ahora esta posición epifenomenalista parece poseer una virtud importante: puesto que las propiedades mentales eran causadas solamente por propiedades fisiológicas pero no figuraban en la cadena de los eventos fisiológicos, la psicología podría desarrollarse en este dominio con una autonomía relativa de la fisiología. Sin embargo, esta posición ha atraído recientemente poco interés puesto que convierte en ineficaces a las propiedades mentales.

El dualismo de propiedades ha sido revivido recientemente, aunque de otra guisa, por Kim (1982a; ver también 1978, 1982b). Éste describe la relación entre propiedades mentales y propiedades físicas como una relación sobrevenida (*supervenience*). El concepto de sobrevenir fue desarrollado originalmente para dar cuenta de las relaciones entre propiedades morales y propiedades físicas. Filósofos morales del siglo XX como G. E. Moore y R. M. Haré argumentaron en contra de cualquier definición de las propiedades morales en términos no morales, pero reconocieron que sería ridículo admitir que dos individuos podrían comportarse de la misma manera y en las mismas circunstancias y uno de ellos ser considerado bueno y el otro malo. El principio del sobrevenir se introdujo para bloquear esta posibilidad. Se mantiene entonces que, si dos individuos o actos son iguales en todas sus propiedades físicas, entonces también son iguales en sus propiedades morales. Para Kim, el rasgo atractivo del modelo del sobrevenir es que ofrece un modo de explicar cómo las propiedades mentales de eventos podrían relacionarse con las propiedades físicas de eventos. Él propuso, sin embargo, restringir el concepto clásico de sobrevenir introduciendo el concepto de «sobrevenir fuerte», que mantiene que, si los individuos comparten las mismas propiedades físicas, entonces *tienen* que compartir las mismas propiedades mentales.

Kim ha mantenido que la tesis del sobrevenir evita el problema de convertir a la mente en algo causalmente ineficaz. De acuerdo con este punto de vista, las propiedades mentales tienen todos los efectos causales de las propiedades físicas sobre las que sobrevienen. Para explicar este punto, Kim (1979) ha comparado el sobrevenir de las propiedades mentales sobre las propiedades físicas con el sobrevenir de las propiedades observables ordinarias de los objetos físicos con sus microestructuras físicas. La microestructura determina la conducta causal de un objeto, pero podemos igualmente atribuir la causalidad a la microestructura y a las propiedades observables. Del mismo modo, las propiedades mentales sobrevenidas tienen todas [119-120] ellas las propiedades causales asociadas con sus propiedades físicas subyacentes. Así, mediante la teoría del sobrevenir podemos reconocer la diferencia entre propiedades mentales y propiedades físicas, hacer un lugar a la eficacia causal de las propiedades mentales y no tener que explicar la interacción entre lo mental y lo físico.

El género más común de objeción que se ha planteado en contra del dualismo, ya sea de objetos o de propiedades, es que resulta metafísicamente extravagante. Se interpreta como violando *la navaja de Occam*, el principio de que debemos ser parsimoniosos en nuestras suposiciones ontológicas y postular solamente aquellas entidades necesarias para nuestra ciencia. Si podernos dar cuenta de todos los fenómenos sin postular entidades o propiedades mentales adicionales, deberíamos hacerlo así. Una razón para adherirse a la navaja de Occam con respecto a la mente es que, si la mente o las propiedades mentales son tan radicalmente diferentes de los objetos o propiedades físicas, entonces podemos pasarlo mal estudiándolas por medio de la ciencia natural. Las técnicas de la investigación científica, incluyendo

las de la ciencia cognitiva, suponen generalmente que estamos tratando con mecanismos físicos que funcionan de acuerdo con principios físicos ordinarios. Por esta razón hasta Popper está de acuerdo en que la investigación debería fundamentarse en supuestos fisicalistas. Popper presenta el dualismo como una posición que estaremos llevados a aceptar como resultado de los fallos de la investigación física a la hora de explicar los fenómenos mentales, no como una posición que debería guiar nuestra investigación. Dada esta aparente falta de resultados del dualismo como un fundamento para la ciencia, necesitamos empezar a considerar las distintas teorías no dualistas que han sido propuestas para reemplazarlo.

5.3 CONDUCTISMO FILOSÓFICO

Una de las primeras alternativas al dualismo que fue cuidadosamente elaborada es la posición conocida como *conductismo filosófico*. Fue bastante popular durante casi el mismo período en el que el conductismo psicológico dominaba la psicología. Aunque el conductismo filosófico y el conductismo psicológico se dan la mano en el rechazo del dualismo, *conductismo* significa algo completamente diferente para los proponentes de esas dos posiciones. Para los psicólogos el conductismo es un programa de investigación empírica que intenta descubrir las leyes que explican la conducta de los humanos y de otros organismos en términos de estímulos ocurrentes y de la historia pasada del condicionamiento del organismo. Su carácter distintivo [120-121] es que rechaza la apelación a eventos mentales para explicar la conducta. Mientras que el conductismo psicológico es un programa de investigación empírico, el conductismo filosófico se interesa primariamente por la semántica de nuestro vocabulario mentalista común. Busca explicar el significado de los términos mentales como *creencia* sin tener que tratarlos como haciendo referencia a substancia alguna. La meta es traducir términos que intentan referirse a actividad mental en términos que hablan solamente de conductas o propensidades a comportarse de ciertas maneras. Así, el conductista filosófico no elimina el discurso mental, pero ofrece una manera de legitimarlo. A pesar de esos diferentes objetivos, los conductistas filosóficos y los conductistas psicológicos se han considerado a sí mismos a menudo como aliados. Skinner (1955), por ejemplo, ha ofrecido análisis conductistas de términos mentales. El conductismo filosófico y el conductismo psicológico se han aliado especialmente al rechazar el punto de vista (central al cognitivismo) de que los eventos mentales son procesos internos a la mente que causa la conducta. En esta sección me centro en la posición del conductismo filosófico y hago notar las similitudes entre él y el conductismo psicológico.

El conductismo filosófico retrotrae sus orígenes a dos amplios movimientos filosóficos discutidos en el capítulo 2. Uno era el Positivismo Lógico, que proponía explicar el significado de las oraciones usadas en una ciencia en términos de las condiciones que verificarían su Verdad. Una de las metas de los positivistas era unificar toda la ciencia. Ellos proponían que, si podemos reducir la discusión de los fenómenos mentales a la discusión de la conducta y de las propensidades a comportarse, obtendríamos el significado de los términos mentales y, a la vez, daríamos el primer paso hacia la unificación de la psicología y la física. La tarea que quedaría entonces sería la de reducir la discusión de la conducta a teorías más básicas de la ciencia física.

El segundo movimiento filosófico que dio lugar al conductismo filosófico fue el análisis de Wittgenstein del lenguaje ordinario. Wittgenstein interpretó muchos problemas filosóficos, tales como el

problema mente-cuerpo, como un resultado de la confusión lingüística y propuso desembarazarse de tales confusiones prestando atención cuidadosa a los modos en que nuestro lenguaje, incluidas nuestras expresiones idiomáticas mentales, se usa en el discurso ordinario.

El *locus classicus* del conductismo filosófico es la monografía de Gibert Ryle, publicada en 1949, *El concepto de lo mental*. En esta obra Ryle presenta el conductismo filosófico no solamente como una alternativa a los puntos de vista tradicionales del dualismo y el materialismo, sino como separándose completamente de la cuestión de la [121-122] relación entre la mente y el cuerpo que él caracterizó como el problema de «el fantasma en la máquina». Ryle caracterizó el problema mente-cuerpo como un resultado de lo que denominaba un «error categorial» puesto que «representa los hechos de la vida mental como perteneciendo a un tipo o categoría (o a un rango de tipos o categorías) lógica» cuando en realidad pertenecen a otro» (1949, p. 16). Ryle usó un ejemplo para explicar la noción de error categorial. Imaginemos a una persona a la que, una vez que se le han enseñado los edificios, facultades, etc., de una universidad, pide que se le enseñe la universidad. La persona en cuestión supone que hay otra entidad comparable a la que ya se le ha enseñado. Puesto que el término *universidad* no se refiere a elementos de la misma categoría que los términos *edificio* y *facultad*, la persona comete un error categorial al buscar la universidad que considera que es algo del mismo género. Similarmente, Ryle afirmaba que se comete un error categorial cuando buscamos la mente como un componente separado del cuerpo adicionalmente a sus diversas partes físicas, o cuando intentamos identificar la mente con alguna parte física del cuerpo.

La alternativa, de acuerdo con Ryle, es reconocer que los vocabularios mentales y físicos pertenecen a tipos lógicos diferentes y siguen reglas diferentes. El vocabulario mental, de acuerdo con Ryle, no intenta describir la conducta de ninguna manera parecida a como el vocabulario fisiológico describe los procesos que ocurren dentro de la gente. Más bien usamos vocabulario mental, de acuerdo con Ryle, para hablar de cómo alguien se comporta o es probable que se comporte. Ryle ilustró esto considerando una serie de expresiones idiomáticas mentales y mostrando cómo pueden acomodarse dentro del enfoque general que él bosqueja. Por ejemplo, podemos explicar lo que queremos decir cuando decimos que alguien cree que lloverá señalando diversas propensiones de conducta, tales como la propensión a llevar un paraguas, a cancelar los planes para una comida campestre y cosas por el estilo[3].

Wittgenstein (1953) y las interpretaciones de Wittgenstein por parte de Malcoln (ver Malcoln, 1984) representan desarrollos posteriores del conductismo filosófico. Al igual que Ryle, Wittgenstein y Malcoln retrotraen el punto de vista mantenido comúnmente de que la mente tiene que ser una entidad especial a la propensión de los filósofos y de otras personas a usar de mala manera el lenguaje ordinario. [122-123] El remedio para esto es el análisis cuidadoso del modo en que funciona el lenguaje ordinario. Una manera en que usamos mal el lenguaje es cuando tratamos los términos mentales como si se refiriesen a eventos que, a continuación, mantenemos que son, por definición, privados (p. ej., dolores o creencias). Nuestra capacidad para usar el lenguaje depende de nuestro usarlo intersubjetivamente. Cuando se usa intersubjetivamente, otras personas pueden averiguar si un hablante particular está usándolo correctamente. Este test de corrección se perdería si las expresiones idiomáticas mentales se refiriesen realmente a eventos privados. (Para un ataque reciente a este argumento, ver Chomsky, 1986.)

Wittgenstein y sus seguidores han mantenido también que podemos descubrir algunas de las constricciones sobre el uso adecuado de los términos mentales atendiendo al modo en que se aprenden. Un dualista podría mantener que aprendemos los términos como creencia y esperanza reconociendo

primero mediante introspección los estados dentro de nosotros que corresponden a creer algo o esperar por algo y, a continuación, aprendiendo a aplicar etiquetas apropiadas a esos estados. Los conductistas filosóficos se cuestionan cómo podríamos enseñar a otra persona a conectar un término con un estado que sólo puede experimentar esa persona. Carecemos de cualquier manera de comprobar y ver si la persona aplicó el término correctamente. La alternativa que ellos proponen es que los términos mentales, como dolor, se aprenden en un contexto público donde, por ejemplo, vemos gente que se queja. Son estos fenómenos públicos los que proporcionan criterios para el uso correcto del vocabulario mental[4].

El conductista filosófico rechaza también el punto de vista de que los términos mentales caracterizan estados de la persona que poseen eficacia causal (p. ej., que nosotros hacemos cosas a causa de nuestras creencias). Términos mentales tales como *creencia* caracterizan disposiciones y, de acuerdo con Ryle (1949), «poseer una propiedad disposicional no es estar en un estado particular, o sufrir un cambio» (p. 43). Por ejemplo, cuando atribuimos fragilidad a un objeto, no estamos afirmando que está en un estado interno particular que [123-124] causa el que se rompa, sino que estamos diciendo solamente que se rompería fácilmente. Del mismo modo, al atribuir una creencia a alguien no estamos haciendo una afirmación sobre los estados internos de la persona, sino simplemente caracterizando a la persona en términos de lo que ella podría hacer en circunstancias particulares. El conductista filosófico afirma que es erróneo tratar a los estados mentales como causas de la conducta. No podemos identificar los estados mentales independientemente de los estados de conducta y de este modo, no podemos tratarlos como causas de la conducta (ver Malcoln, 1984).

El conductismo filosófico, al rechazar los estados mentales internos, es claramente incompatible con el programa de la ciencia cognitiva consistente en explicar la conducta en términos de modelos de procesamiento. Para Ryle, hablar de procesamiento interno no añade nada a lo que entendemos sobre una persona cuando sabemos sus propensiones a comportarse de maneras específicas. Para Wittgenstein, la psicología experimental es un esfuerzo descaminado para introducir el habla psicológica dentro de la ciencia experimental. Su propuesta es que en lugar de eso deberíamos intentar entender los fenómenos psicológicos examinando cómo ha evolucionado el lenguaje para habérselas con la conducta humana.

Al igual que el conductismo psicológico, el conductismo filosófico ha perdido popularidad en los últimos años. Esto se debe en gran medida al reconocimiento de dificultades aparentemente serias en esta posición. Es obvio que no podemos simplemente traducir términos mentales en descripciones de conductas, puesto que los estados mentales como las creencias no siempre se manifiestan en conducta. Los conductistas filosóficos han intentado hacer equivalentes los términos mentales a términos que adscriben *disposiciones* o *propensiones* a comportarse de ciertas maneras bajo estímulos apropiados. Por ejemplo, mi creencia de que tengo una cita a las 10 en punto de la mañana podría identificarse no con alguna conducta que estoy llevando a cabo ahora, sino con las propensiones que tengo a comportarme de maneras particulares. Por ejemplo, si tengo esta creencia, entonces, si observo que mi reloj señala las 9.59, me levantaré de repente y saldré como un rayo de mi despacho.

El análisis disposicional no evita, sin embargo, todos los problemas. En primer lugar, los estados mentales individuales no pueden generalmente hacerse equivalentes con distintas disposiciones a comportarse. Mi creencia de que tengo una cita a las 10 en punto de la mañana estará asociada con una gran variedad de disposiciones. De hecho, este conjunto puede ser ilimitado e incluirá un buen número de disposiciones que difícilmente consideraríamos hasta el [124-125] mismo momento en que surgen. Por

ejemplo, si estoy retenido en el despacho del decano a las 9.59, puede que no me levante y salga como un rayo, sino que puede que pida hacer una llamada telefónica[5]. La variedad de tales posibilidades parece no tener fin. El conductista filosófico parece que está comprometido a analizar creencias en términos de largas listas potencialmente infinitas de oraciones condicionales que introducen nuevos problemas. Una de las presuntas virtudes del conductismo filosófico era dar cuenta de cómo aprendemos a usar los términos mentales mediante la experiencia. Sin embargo, la propuesta de que los términos mentales han de hacerse equivalentes con listas potencialmente infinitas de enunciados condicionales convierte en dudosa esa propuesta, puesto que tendríamos que aprender esta lista potencialmente infinita para aprender los términos mentales.

Hay un segundo problema que es más serio. Las oraciones condicionales que se supone que dan las equivalencias de significado de los términos mentales emplean ellas mismas casi inevitablemente términos mentales. En el ejemplo de mi creencia de que tengo una cita a las 10 en punto de la mañana, he usado una oración condicional sobre lo que sucedería *si me doy cuenta* de la hora que marca mi reloj. El término me doy cuenta es también un término mental, al que se le debe dar a su vez una traducción en oraciones condicionales. Esto sugiere que estamos atrapados en un círculo de términos mentales en el que los correlatos conductistas de un término pueden enunciarse solamente usando otros términos mentales. Los críticos han argumentado que jamás podemos salir fuera de este círculo, puesto que todas las pretendidas traducciones de los términos mentales emplearían ellas mismas términos mentales. (Ver Chisholm, 1957; Geach, 1957.)

Un tercer problema tiene que ver con los modos en los que podríamos asignar disposiciones a los agentes. No podemos describir disposiciones excepto sobre la base de conducta ya realizada. Pero, como objeta Armstrong (1968), la conducta previa subdetermina siempre las disposiciones. Podemos siempre imputar una gran variedad de disposiciones para dar cuenta de cualquier conducta particular. Si consideramos que los términos mentales adscriben disposiciones particulares a agentes, entonces tenemos que suponer que hay [125-126] algo sobre los agentes que es lo que fija qué disposición ha de adscribirse. Esto solamente parece posible si tratamos los términos mentales como refiriéndose a estados internos determinados cuyo carácter fija la disposición incluida. Esto, sin embargo, viola las restricciones impuestas por el conductismo filosófico.

Uno de los fundamentos sobre los que fue construido el conductismo filosófico, la teoría verificacionista del significado, ha sido también desafiada en los años recientes. Quine (1953/1961a) criticó como un dogma del empirismo la suposición de que podríamos definir lógicamente los términos teóricos de manera observacional, y cada vez más filósofos de la ciencia han llegado a reconocer que podríamos tener que aceptar términos de nuestro vocabulario científico que no pueden reducirse lógicamente a términos observacionales. Si abandonamos el verificacionismo en general, parecería que no hay razón para no hacerlo así también en el caso del discurso mental. El hacer esto permite que los términos mentales se introduzcan dentro del discurso psicológico de la misma manera que los términos teóricos se introducen en una ciencia (ver Fodor, 1968; Geach, 1957; Sellars, 1963b). Esto, desde luego, deja abierta la cuestión de a qué se refieren esos términos teóricos. En el capítulo 6 discuto algunos intentos alternativos de explicar la referencia de los términos mentales dentro de una armazón fisicalista.

5.4 RESUMEN INTERMEDIO DEL PROBLEMA MENTE-CUERPO

En este capítulo he examinado dos puntos de vista filosóficos sobre las relaciones de la mente y el cerebro que han sido muy influyentes en la configuración de las discusiones sobre este problema. Descartes diferenció la mente y el cerebro, y él y otros han intentado mostrar en qué aspectos la mente es un género diferente de entidad del de los objetos físicos como el cerebro. El dualismo se ha encontrado con un gran número de problemas al explicar la relación entre la mente y el cuerpo y ha sido acusado de inflar nuestra ontología de modo innecesario. El conductismo filosófico evita el dualismo negando que los estados mentales sean estados internos de las personas. En lugar de esto intenta analizar los estados mentales en términos de disposiciones de conducta, una jugada que se encuentra con un gran número de problemas. Ambos puntos de vista se han enfrentado así a severas críticas de modo que, mientras que algunos filósofos aún los mantienen, muchos han ensayado otras opciones. Algunas de ellas se consideran en el capítulo siguiente.

6. EL PROBLEMA MENTE-CUERPO: VERSIONES DEL MATERIALISMO

6.1 INTRODUCCIÓN

En el capítulo anterior introduje el problema mente-cuerpo y discutí dos respuestas filosóficas al mismo. Otra respuesta tradicional mantiene que los estados mentales son estados del cerebro. Este punto de vista, que comúnmente recibe los nombres de *materialismo* y *fisicalismo*, puede remontarse al menos hasta Hobbes y fue desarrollado adicionalmente por Gassendi y LaMettrie en los siglos XVll y XVlll. La mayor parte de los filósofos contemporáneos y probablemente la mayor parte de los científicos cognitivos apoyan el materialismo. Sin embargo, desde 1950 los filósofos han intentado enunciar de manera más precisa la tesis del materialismo. Examino en este capítulo tres versiones contemporáneas, cada una de las cuales tiene un conjunto diferente de consecuencias para la ciencia cognitiva.

6.2 LA TEORÍA DE LA IDENTIDAD COMO TIPO MENTE-CEREBRO

La expresión «Teoría de la Identidad» se refiere propiamente hablando al enfoque desarrollado en los años cincuenta por U.T. Place (1966/1970), Herbert Feigl (1958/1967, 1960/1970) y J. J. C. Smart (1959/1971), que fue adoptado por un buen número de filósofos en la década siguiente. Esas teorías proponían que los estados mentales eran idénticos a estados del cerebro. La expresión puntualizadora «tipo» ha sido introducida más recientemente para distinguir este punto de vista de otro más débil que alcanzó cierta preeminencia en los setenta y en los ochenta, que es conocido como «la Teoría de la Identidad como Instancia» y que se discute en una sección posterior. La distinción tipo/instancia se refiere a la diferencia entre una clase de eventos (el tipo) y un miembro específico de la clase (una instancia). El término *silla* identifica un tipo de objeto, mientras que la silla de mi despacho es una instancia particular de ese tipo. La Teoría de la Identidad como Tipo mantiene que todas las instancias de un tipo particular de estado mental (p. ej., tener la experiencia de un cierto [127-128] género de dolor o ver un cierto color) son idénticas a instancias de un tipo de evento neural correlacionado (p. ej., un cierto modelo de excitaciones neurales).

Una de las inspiraciones principales de la Teoría de la Identidad como Tipo fue la obra de neurofisiólogos como Köhler, Penfield y Hebb, que se consideró que señalaba un isomorfismo entre informes fenoménicos y neuroprocesos específicos. Feigl interpretó que la tarea de los filósofos era proporcionar una «clarificación lógica y epistemológica de los conceptos por medio de los cuales podemos formular y/o interpretar esas correlaciones» (1960/1970, p. 35). Los puntos de vista epifenomenalistas tales como los que se han discutido en el capítulo anterior proporcionan una manera de interpretar esos resultados. De acuerdo con el epifenomenalismo, el análisis causal completo de la conducta se centrará en las interacciones entre los eventos cerebrales, pero habrá un segundo conjunto de relaciones causales de acuerdo con el cual algunos estados cerebrales producirán estados fenoménicos.

Feigl rechaza el epifenomenalismo, caracterizando su tratamiento de los estados mentales como la postulación de «"guirnaldas" puramente mentales», que él considera como una «solución muy extraña»: «Esas leyes de correspondencia son peculiares en que puede decirse que postulan "efectos" (estados mentales como variables dependientes) que no funcionan por sí mismos como "causas" (variables independientes), o que al menos no parecen necesitarse como tales, para cualquier conducta observable» (p. 37). La alternativa que avanzó Feigl es que los estados mentales se refieren a exactamente los mismos estados que los términos físicos incluso si describen los estados de manera diferente: «Utilizando la distinción de Frege entre *Sinn* ("significado", "sentido", "intensión") y *Bedeutung* ("referente", "*denotación*", "extensión"), podemos decir que los términos neurofisiológicos y los correspondientes términos fenoménicos, aunque se diferencian ampliamente en sentido y, por tanto, en los modos de confirmación de los enunciados que los contienen, tienen *referentes* idénticos» (p. 38). Los teóricos de la identidad invocan entonces el análisis de Frege de los enunciados de identidad (ver capítulo 2) para explicar cómo pueden ser idénticos los estados mentales y los estados físicos: las expresiones idiomáticas mentales y las expresiones idiomáticas físicas son descripciones diferentes de los mismos estados.

Un problema discutido en los primeros escritos sobre la Teoría de la Identidad era el rango de los estados mentales a los que se aplica este enfoque. Place fue el primero en proponer la Teoría de la Identidad, aunque aceptó la identificación del conductista filosófico de algunos estados mentales con disposiciones. El mantenía solamente [128-129] que algunos otros conceptos mentales podrían no referirse a disposiciones: mantuvo que había aquí «un residuo intratable de conceptos arracimados sobre las nociones de conciencia, experiencia, sensación, imágenes mentales, donde es inevitable alguna suerte de historieta sobre procesos internos» (1956/1970, p. 43; ver también 1988). Esos procesos internos serían procesos en el cerebro.

Sin embargo, otros proponentes de la Teoría de la Identidad la generalizaron hasta el punto de mantener que todos los términos mentales, incluyendo los que el conductista filosófico había analizado como refiriéndose a disposiciones, se referían realmente a estados cerebrales. La extensión era muy natural. En otras disciplinas los enunciados de disposición se reducen a menudo a enunciados sobre la constitución interna del objeto que posee la disposición. Por ejemplo, la fragilidad del cristal se identifica con su estructura física. Similarmente, los teóricos de la identidad propusieron que era el estado del cerebro el que daba cuenta de que una persona esté en determinado estado mental tal como tener una creencia particular (ver Armstrong, 1968).

El problema más difícil al que se enfrentaban los primeros proponentes de la Teoría de la Identidad consistía en clarificar lo que significaba la afirmación de que los estados mentales eran idénticos a estados cerebrales. El Teórico de la Identidad está comprometido con lo que Smart (1959/1971) ha denominado la identidad en el «sentido estricto», no con la mera correlación. (Popper, como he discutido en el capítulo anterior, ha malinterpretado la posición del Teórico de la Identidad tratándola como si fuera correlación.) Muchos críticos han encontrado que la idea de una identidad estricta de los estados mentales y físicos es o ininteligible u obviamente falsa, puesto que los términos físicos y los términos mentales difieren de modo muy importante en sus significados. La objeción siguiente es bastante típica:

> Decir que la conciencia es una forma de materia o de movimiento es usar palabras sin

significado. La identificación de la conciencia y del movimiento puede, de hecho, no ser refutada jamás; pero sólo porque aquel que no ve el absurdo de tal enunciado jamás puede hacer ver nada... Si él no puede ver que, aunque la conciencia y el movimiento estén *relacionados* tan íntimamente como se quiera, *queremos decir* cosas diferentes mediante las dos palabras; que, aunque la conciencia pueda estar *causada* por el movimiento, esto no es lo que queremos decir mediante movimiento en igual medida que no es eso lo que queremos decir mediante queso verde: si él no puede ver esto no hay modo de discutir con él. (Pratt. 1922/1957, p. 266.)

Las objeciones a la Teoría de la Identidad se presentan a menudo en términos de la *Ley de Leibniz*, que, como vimos en el capítulo 2, mantiene que, si dos términos se refieren al mismo objeto, entonces [129-130] cualquier propiedad que se predica verdaderamente del objeto al que se hace referencia por el primer término tiene que ser también predicada verdaderamente del objeto cuando se hace referencia a él mediante el segundo término, y viceversa. Los críticos afirman encontrar un buen número de propiedades que podrían atribuirse bien a eventos físicos o bien a eventos mentales, pero no a ambos. Una de tales propiedades es la intencionalidad (ver capítulos 3 y 4), que se piensa que se aplica a los eventos mentales y no a los eventos físicos. Si es verdad que los eventos mentales exhiben intencionalidad y los eventos cerebrales no, entonces los eventos mentales y los eventos cerebrales no son idénticos. Shaffer (1965) plantea esta objeción:

Cuando informo de que repentinamente recuerdo que Henry estaba enfermo, la intencionalidad de este informe, esto es: que es sobre Henry y sobre su enfermedad, es una parte esencial de él. Este rasgo intencional se pierde si informamos de que un evento neural particular ha ocurrido de manera repentina: tal informe no sería en absoluto sobre Henry, sino sólo sobre un evento cerebral. Desde luego, podríamos dar siempre esas funciones nuevas a los eventos cerebrales, pero esto sería redefinir expresiones fisicalistas en vez de redefinir expresiones mentalistas, dejándonos donde habíamos comenzado (p. 95).

Hay otras propiedades que parecen comportarse de modo similar. Por ejemplo, cuando experimentamos una post-imagen, parecemos experimentar algo con un color y una forma particulares. Puesto que no hay ningún objeto con ese color y esa forma que vemos efectivamente, es común decir que el objeto existe en nuestra mente. Pero no podríamos decir que existía en nuestro cerebro un objeto de esa forma y de ese color. Por consiguiente, hay objetos en la mente que no están en el cerebro.

Los eventos físicos tienen también propiedades de las que parecen carecer los eventos mentales. Por ejemplo, todos los eventos físicos tienen coordenadas espaciales —ocurren en algún lugar—. Pero, como Shaffer (1965) afirma:

por lo que respecta a los pensamientos, no tiene sentido hablar de que un pensamiento está localizado en algún lugar o lugares del cuerpo. Si informo de que he pensado algo repentinamente, la cuestión de en qué lugar de mi cuerpo ha ocurrido ese pensamiento sería un sinsentido completo (p. 97).

De este modo, concluyen Shaffer y otros, los eventos mentales no pueden ser eventos cerebrales.

Otra objeción común a la Teoría de la Identidad mantiene que los eventos mentales y los eventos físicos no pueden ser lo mismo puesto [130-131] que estamos familiarizados con ellos de maneras diferentes. Se afirma que somos conscientes directamente de los estados mentales —no necesitamos realizar investigaciones para averiguar cosas acerca de ellos—. Tenemos lo que se llama *acceso privilegiado* a nuestra vida mental. Sin embargo, sólo podemos averiguar cosas sobre los estados de nuestros cerebros muy indirectamente, si es que podemos. Puesto que tenemos acceso privilegiado a nuestros estados mentales pero carecemos de tal acceso privilegiado a nuestros estados cerebrales, los críticos mantienen que ambos no pueden ser lo mismo.

El artículo clásico de Smart (1959/1971) en defensa de la Teoría de la Identidad consiste en gran medida en intentos de rechazar las objeciones de esta clase clarificando lo que está incluido en una afirmación de identidad. Para empezar, mantiene él, las afirmaciones de identidad no son afirmaciones de necesidad lógica que puedan establecerse analizando cómo usamos el lenguaje. Más bien son afirmaciones contingentes que podrían ser falsas. El Teórico de la Identidad quiere contemplar la posibilidad de que los eventos mentales puedan ser algo distinto de los eventos cerebrales, pero pretende que en nosotros son estados cerebrales. Por consiguiente, las objeciones de que los términos mentales y los términos físicos tienen significados diferentes no cuentan en contra de la tesis de la identidad. Smart se opone a la afirmación de que mucha gente no sabe nada acerca de sus procesos cerebrales, mientras que sí que saben sobre sus estados fenoménicos, afirmando que la Teoría de la Identidad no depende de cómo entiende la gente los conceptos usados para expresar la afirmación, sino solamente de si ambos términos se refieren de hecho a la misma cosa. Él afirma que puede haber enunciados contingentes de la forma «A es idéntico a B», y una persona puede saber perfectamente que algo es un A sin saber que es un B. «Un campesino iletrado podría perfectamente ser capaz de hablar sobre sus sensaciones sin saber nada sobre sus procesos cerebrales, lo mismo que puede hablar sobre iluminación aunque no sepa nada sobre electricidad» (Smart, 1959/1971, p. 58).

Como respuesta a la objeción de que la mayor parte de la gente adscribe propiedades diferentes a experiencias mentales y a experiencias físicas, Smart mantiene que esto es simplemente un rasgo del uso de nuestro lenguaje diario. En el futuro podríamos revisar nuestro lenguaje para permitir, por ejemplo, predicaciones de intencionalidad a los estados cerebrales. De hecho, el propio Smart ha defendido una revisión de nuestro lenguaje. Para hacer frente a la objeción de que nuestro discurso fenoménico parece referirse a propiedades fenoménicas (p. ej., propiedades de color) que son distintas de las [131-132] propiedades físicas, Smart ha propuesto lo que él denomina una terminología «neutral respecto al tópico». Así, ha recomendado traducir «Veo una post-imagen de un naranja amarillento» como «*Sucede algo que es semejante a lo que sucede cuando* tengo mis ojos abiertos, estoy despierto, y hay una naranja bien iluminada frente a mí, esto es: cuando realmente veo una naranja» (Smart, 1959/1971, p. 61). El objeto de traducir informes a la forma neutral respecto al tópico es evitar la suposición de que esos informes lo son sobre propiedades peculiarmente mentales que podrían no identificarse con propiedades físicas. La propuesta de Smart trata también de la objeción de la post-imagen. El habla de post-imágenes sugiere que en la mente hay un objeto que corresponde a la imagen, pero la versión tópiconeutral de Smart elimina cualquier tentación de decir que está presente un objeto fenoménico cuando vemos post-imágenes. Más bien nos lleva a decir que lo que está ocurriendo son simplemente eventos semejantes a

los que ocurren cuando vemos objetos reales, externos. La propuesta de Smart de las traducciones tópico-neutrales ha sido objeto de controversia. Para muestras de algunas críticas, ver Cornman (1962/1971) y Margolis (1978).

Como he observado anteriormente, muchas de las objeciones a la Teoría de la Identidad han descansado sobre la Ley de Leibniz. Implícitamente, lo que Smart está haciendo es intentar mostrar que las exigencias de la Ley de Leibniz pueden cumplirse de modo efectivo mediante maniobras lingüísticas apropiadas. Otros defensores de la Teoría de la Identidad han adoptado una estrategia diferente que niega la aplicabilidad de la Ley de Leibniz a esos contextos. Cornman (1962), por ejemplo, mantiene que la Ley de Leibniz no se viola cuando se encuentra que los predicados mentales no son aplicables a estados físicos, o viceversa. Solamente tendríamos una violación si una predicación tuviese un valor de verdad diferente que otra. Pero en este caso la predicación inaplicable no es ni verdadera ni falsa. El consideró que esto mostraba que estamos ante un caso de error categorial tal como Ryle lo había descrito. Cornman, sin embargo, extrajo una moraleja diferente de la de Ryle. Mantuvo que es legítimo postular identidades categoriales cruzadas y que en tales casos la Ley de Leibniz es simplemente inaplicable. Apoyó su análisis considerando otro caso:

> Decimos que la temperatura de un gas es idéntica a la energía cinética media de las moléculas del gas. Pero, aunque podemos decir que la temperatura de un determinado gas es de 80 grados centígrados, es seguramente un error en cierto sentido el decir que la energía cinética media de las moléculas del gas es 80 grados centígrados. Si este error es lo que he llamado un error categorial, entonces esto es un caso de identidad categorial cruzada. [132-133] Si también es un error categorial hablar de un proceso cerebral desvaneciente o débil, entonces tenemos algún fundamento para pensar que la identidad de mente y cuerpo no sería una identidad categorial cruzada y, por tanto, que la Teoría de la Identidad no necesita involucrar dificultades conceptuales. (Cornman, 1962/1971.)

Como se acaba de poner de manifiesto anteriormente, los defensores de la Teoría de la Identidad han interpretado la identidad de los eventos mentales y los eventos físicos como algo que es verdadero pero que podría haber sido falso. A tales enunciados se hace referencia como *contingentes*. Basándose en su análisis de los enunciados modales (ver capítulo 2), Saúl Kripke (1972) ha argumentado que las identidades contingentes son imposibles. Como hemos visto, Kripke mantenía que los enunciados necesarios son verdaderos en todos los mundos posibles, y que un *designador rígido* es un término que selecciona la misma entidad en cualquier mundo posible en el que la entidad existe. Un designador *no rígido* es un término que cambia su referente a través de los mundos posibles. (Por ejemplo, «Jimmy Cárter» es un designador rígido. Selecciona la misma persona en cualquier mundo en el que existe Cárter. El término «El Presidente número 39 de los Estados Unidos» no es, sin embargo, un designador rígido, puesto que podría haber sido elegida otra persona en 1976.) Kripke argumentó que los enunciados de identidad propiamente dichos tienen que poner en equivalencia términos que son designadores rígidos. Esto entraña que todos los enunciados de identidad son necesarios y no contingentes, puesto que ambos nombres seleccionarán el mismo objeto en cada mundo posible. Una vez que ha intepretado como necesarias todas las afirmaciones de identidad, Kripke argumenta que los estados mentales no pueden ser

idénticos a los estados físicos. Mantuvo que los términos que se refieren a estados mentales y a estados cerebrales son designadores rígidos. Puesto que podemos estipular un mundo posible en el que los términos que se refieren a los estados mentales no se refieren a las mismas cosas que los términos que se refieren a estados cerebrales, esos designadores rígidos no pueden seleccionar los mismos objetos. Por consiguiente, no pueden estar en relaciones de identidad.

Aunque los argumentos de Kripke son sofisticados, muchos filósofos y, probablemente, muchos investigadores empíricos encuentran que no van al grano a la hora de habérselas con cuestiones empíricas. Parte de la dificultad surge de la cuestión de cómo determinamos cuáles son los mundos posibles. La respuesta de Kripke, como hemos visto antes, es que estipulamos los mundos posibles: determinamos qué rasgos del mundo presente vamos a alterar para llegar al mundo posible. Este tratamiento de los mundos posibles tiene, [133-134] sin embargo, la infortunada consecuencia de hacer que la evaluación de afirmaciones de lo que es posible dependa de nuestra capacidad de imaginar ciertas situaciones. Pero está claro que podemos pensar que algo es posible y descubrir más tarde que no lo es. La gente pensaba que la Estrella de la Tarde podría dejar de existir aunque continuase existiendo la Estrella de la Mañana, pero ahora sabemos que esto no es posible. Similarmente, aunque podemos concebir los estados cerebrales como existiendo sin estados mentales concomitantes, esto podría no ser posible efectivamente. La legislación lingüística no puede decidir este problema. Por tanto, incluso si concedemos a Kripke la afirmación de que todas las identidades tienen que ser identidades necesarias, no se sigue el rechazo de la Teoría de la Identidad. (Para otras críticas filosóficas de los argumentos de Kripke, ver Barnette, 1977; Feldman, 1974, 1980; Kirk, 1982; Lycan, 1974; Maxwell, 1978 y Sher, 1977.)

He observado al comienzo de esta discusión que los proponentes de la tesis de la identidad se contemplan a sí mismos como si estuvieran dando una exposición lógica de avances de investigación en neurociencia. Pero la investigación en neurociencia, como han señalado muchos críticos, jamás podría establecer otra cosa que una correlación entre eventos mentales y eventos cerebrales. El que adoptemos una afirmación de correlación (que incluso los dualistas pueden aceptar) o una afirmación de identidad parece ser un problema que va más allá de la evidencia empírica. Los proponentes de la Teoría de la Identidad apelan a menudo a la navaja de Occam para apoyar su posición. La navaja de Occam nos invita a aceptar una teoría que postula menos entidades en lugar de una que postula más entidades sin que haya ganancia en poder explicativo. Feigl estaba usando implícitamente la navaja de Occam en el pasaje citado anteriormente en el que comentaba el carácter peculiar del epifenomenalismo. Smart (1959/1971) se refería a ello directamente en su defensa de la Teoría de la Identidad:

> ¿Por qué queremos resistimos [al paralelismo]? Principalmente debido a la navaja de Occam. Me parece que la ciencia está dándonos cada vez más un punto de vista donde los organismos son capaces de ser vistos como mecanismos psicoquímicos: parece que incluso la conducta del hombre mismo será explicable algún día en términos mecánicos. No parece haber nada en el mundo, por lo que respecta a la ciencia, excepto disposiciones cada vez más complejas de constituyentes físicos. Todo excepto en un lugar: la conciencia... Que todo sea explicable en términos de la física, excepto la ocurrencia de sensaciones, me parece francamente increíble (p. 54).

Sin embargo, los críticos de la Teoría de la Identidad objetan que [134-135] en este caso no es posible arreglárnoslas con menos entidades. Las propiedades mentales y las propiedades físicas nos parecen diferentes y necesitamos explicar esta diferencia. Esto exige postular al menos propiedades duales, si no objetos duales.

Los debates entre los teóricos de la identidad y sus críticos parecen quedar en tablas: ninguna de las partes es capaz de convencer a la otra. (Para una discusion adicional de la Teoría de la Identidad como Tipo, ver Enc, 1983; Hill, 1984.) Dennet (1979) ha comentado cómo este problema polariza a las personas:

> La afirmación definitoria de la Teoría de la Identidad de que los eventos mentales no son meramente paralelos a, coincidentes con, causados por, o acompañamientos de, eventos cerebrales, sino que son (estrictamente idénticos a) eventos cerebrales, divide a la gente de una manera curiosa. Para algunas personas parece obviamente verdadero (aunque puede haber un pequeño lío con los detalles a la hora de expresarlo apropiadamente), y para otros parece, con la misma fuerza, obviamente falso. Los primeros tienden a contemplar todos los intentos de resistirse a la Teoría de la Identidad como algo motivado por un temor irracional al avance de las ciencias físicas, una especie de hylefobia humanística, mientras que los últimos tienden a despachar a los teóricos de la identidad motejándolos de ciegos y descaminados adoradores de la ciencia que no se dan cuenta del manifiesto ridículo de la afirmación de identidad (p. 252).

El decidir entre la afirmación de la identidad y el paralelismo puede ser imposible si apelamos solamente a cómo describimos los estados mentales y físicos y las intuiciones de las personas respecto de si un estado cerebral podría poseer propiedades mentales, y viceversa. Un enfoque alternativo es interpretar las afirmaciones de identidad como afirmaciones hechas en el curso de la investigación científica y considerar cómo evalúan típicamente los científicos sus afirmaciones.

Generalmente, las afirmaciones de identidad se hacen al principio de la investigación científica y no al final de la investigación. La Ley de Leibniz no se usa para evaluar la corrección de un enunciado de identidad, sino para generar hipótesis empíricas nuevas que han de ser investigadas. Las afirmaciones de identidad se avanzan a menudo cuando los investigadores piensan que podría haber una identidad entre entidades que previamente se han investigado de manera separada en campos diferentes de investigación. La Ley de Leibniz se convierte en relevante cuando los investigadores intentan usar lo que un campo conoce sobre la entidad para tratar con problemas que surgen originalmente en otro dominio. Por ejemplo, Mendel (1865) postuló inicialmente factores (más tarde llamados *genes*) que él consideró que eran responsables de la herencia de rasgos entre padres e [135-136] hijos. Por otra parte, los cromosomas se identificaron en la investigación citológica, donde los procedimientos involucrados en la meiosis y en la mitosis sugirieron que tenían que desempeñar algún papel importante en la herencia de una célula a otra. Boveri (1905) y Sutton (1903) propusieron, sobre la base de la evidencia de que la distribución anormal de cromosomas llevaba a un desarrollo anormal, que los cromosomas eran las unidades de la herencia. Esto generó la afirmación de identidad de que los factores mendelianos eran unidades en los cromosomas, lo que entonces llevó al programa de investigación extremadamente fructífero de la escuela de Morgan. La información que se conocía sobre los cromosomas se aplicó a los genes, y viceversa. Lo

fructífero de la afirmación de identidad era algo que sólo podría evaluarse como resultado de la investigación que resultó de ella y no en la época en que fue avanzada (ver Bechtel, en prensa b, capítulos 5 y 6; Churchland, 1986; Darden y Maull, 1977; Wimsatt, 1976). Aplicar la misma perspectiva al caso mente-cerebro exigiría tratar la Teoría de la Identidad como una hipótesis de trabajo que ha de ser investigada posteriormente. Si, sobre la base de las afirmaciones de identidad psicofísicas, podemos usar lo que se conoce sobre los eventos mentales para hacer avanzar nuestra comprensión de los procesos neurales y viceversa, entonces estará justificada una afirmación de identidad más bien que una afirmación de correlación.

Al apoyar los estados internos como factores causales que pueden usarse al explicar la conducta, la Teoría de la Identidad es más compatible con las preocupaciones de la ciencia cognitiva actual que lo fue el conductismo filosófico. Pero la Teoría de la Identidad sólo permite la apelación a eventos internos suponiendo que los tipos de eventos mentales son idénticos a los tipos de eventos neurales. Por tanto, las teorías cognitivas se limitan a modos de clasificar eventos mentales que tienen correspondencia biunívoca con los usados en neurociencia. Tal conexión puede arruinar los esfuerzos de los cognitivistas, puesto que el modo más fructífero de clasificar eventos para propósitos cognitivos puede no corresponder al requerido para la neurociencia (ver Fodor, 1974, y p. 146 de este volumen). Además, en la medida en que la Teoría de la Identidad se inspiró en el trabajo de la neurociencia, hay al menos la sugerencia de que las teorías cognitivas deben correr parejas con las teorías de la neurociencia. Así, la Teoría de la Identidad parece dar primacía a las neurociencias sobre la investigación de la ciencia cognitiva. En el mejor de los casos, las teorías cognitivas podrían describir en términos cognitivos los mismos procesos que describe la neurociencia en un vocabulario más físico.

Uno de los tópicos a los que muchos materialistas recientes se han [136-137] opuesto con la Teoría de la Identidad como Tipo ha sido la supuesta correlación de los eventos mentales con los eventos físicos. Esos materialistas, sin embargo, no están de acuerdo sobre la respuesta adecuada. Los *materialistas eliminativos* consideran esto como una razón para eliminar de nuestro lenguaje el habla sobre lo mental a favor del habla sobre el cerebro, mientras que los defensores de la Teoría de la Identidad como Instancia proponen que deberíamos continuar hablando sobre fenómenos mentales pero reconociendo que sólo son los eventos mentales individuales los que pueden identificarse con eventos físicos. Niegan que podamos relacionar tipos de eventos mentales con tipos de eventos físicos. Vuelvo a considerar esas posiciones en las dos secciones siguientes.

6.3 MATERIALISMO ELIMINATIVO

El materialismo eliminativo comienza afirmando que la investigación en neurociencia no demuestra la correlación entre procesos del cerebro y procesos mentales que afirma la Teoría de la Identidad como Tipo y argumenta que esto es una razón para reemplazar el habla sobre lo mental por el habla sobre estados del cerebro. Dicho de manera más exacta: afirman que no hay fenómenos mentales y que los que afirman que los hay están equivocados[1].

En parte los eliminativistas parecen unos materialistas más directos que los Teóricos de la Identidad. Feigl, en una posdata que añadió diez años después de que escribiese un ensayo en el que defendía la Teoría de la Identidad, la repudió a favor del Eliminativismo. Hizo esto dado que concluyó que los

fenómenos mentales no podrían identificarse de manera precisa con actividades cerebrales. Propuso que, en lugar de intentar forzar una integración más estricta entre los conceptos mentales y los conceptos físicos, lo que podríamos hacer era comenzar a usar los conceptos físicos como reemplazos de los conceptos mentales. Predijo que, una vez que la neurociencia se desarrollase suficientemente, no necesitaríamos ya hablar más de otras [137-138] personas que experimentan sentimientos de placer o cosas por el estilo, sino que, en su lugar, usaríamos los nuevos conceptos de la neurociencia (Feigl, 1958/1967, pp. 141-142).

Paul Feyerabend alcanzo la misma conclusión un poco antes. Feyerabend (1963/1970) mantuvo que en la misma formulación de los enunciados de identidad psicofísica, el Teórico de la Identidad parecía estar comprometido con propiedades psicológicas no reducibles. Apoyó básicamente el mismo remedio que Feigl, proponiendo que deberíamos abandonar el lenguaje mentalista lo mismo que hemos abandonado el lenguaje sobre posesiones demoníacas una vez que se ha desarrollado la teoría moderna sobre la epilepsia. Deberíamos reemplazar la terminología mentalista por terminología nueva extraída de la neurociencia. Con el objeto de ilustrar el género de objeción que esperaba, cita la defensa de J.L. Austin (1955-1957/1960) del lenguaje ordinario:

> Nuestro depósito común de palabras incorpora todas las distinciones que los hombres han considerado valioso establecer, y las conexiones que han considerado valioso señalar a lo largo de la vida de muchas generaciones: seguramente éstas son con toda probabilidad las más numerosas, las más correctas, puesto que han aguantado la larga prueba de la supervivencia del más apto, y más sutiles que cualesquiera otras que usted o yo verosímilmente pensemos (p. 182).

Sin embargo, Feyerabend (1963/1970) no resulta impresionado por tales afirmaciones sobre el lenguaje ordinario:

> En primer lugar, tales expresiones idiomáticas [del lenguaje ordinario] están adaptadas a las *creencias* y no a los *hechos*. Si esas creencias se aceptan ampliamente; si están conectadas íntimamente con los temores y las esperanzas de la comunidad en la que ocurren; si se defienden y se refuerzan con la ayuda de poderosas instituciones; si toda la vida de uno se lleva de alguna manera de acuerdo con ellas, entonces el lenguaje que las representa se considerará como el que tiene más éxito. Al mismo tiempo no se ha tocado la cuestión de la verdad de las creencias.
>
> La segunda razón por la que el éxito de una expresión idiomática «común» no está en absoluto al mismo nivel en que está el éxito de una teoría científica reside en el hecho de que el uso de una expresión idiomática tal, *incluso en situaciones observacionales concretas*, difícilmente puede considerarse como un *test*. No hay intento alguno, como lo hay en las ciencias, de conquistar nuevos campos y ensayar la teoría en ellos (p. 144).

Además de este repudio global del *status* privilegiado de nuestra habla sobre lo mental, Feyerabend rechaza también la afirmación de Descartes de que el discurso mental es infalible de tal manera que, si pensamos que estamos en cierto estado mental, ninguna otra evidencia [138-139] podría establecer que no lo

estamos. En contraste con esto, Feyerabend mantiene que los informes de los estados mentales descansan sobre expresiones idiomáticas lingüísticas y que podríamos necesitar revisarlas. Además afirma que nuestras expresiones idiomáticas mentalistas no son neutrales respecto a la teoría, sino que llevan codificada una teoría sobre eventos mentales privados. Aunque esta teoría esté fuertemente arraigada, puede ser errónea. Si lo es, nuestro uso continuado de discurso mentalista no hace otra cosa que perpetuar un mito.

En sus primeros escritos Rorty concordaba con los ataques básicos de la posición de Feyerabend. Sin embargo, de una manera más fuerte que Feyerabend, Rorty se concentró en el punto de conexión entre las antiguas y las nuevas armazones y defendió la identificación de objetos especificados en la vieja armazón con los especificados en la nueva. De este modo apoyó la «forma de desaparición» de la Teoría de la Identidad que mantiene que, a medida en que la ciencia avanza, introducimos un nuevo vocabulario para hablar sobre aquello para lo que previamente usamos otro vocabulario. Cuando lo hacemos así reconocemos que el viejo vocabulario es inadecuado de modo que:

> la relación en cuestión no es estrictamente identidad, sino más bien la suerte de relación que, para decirlo crudamente, se da entre entidades existentes y entidades no existentes cuando la referencia a las últimas ha servido alguna vez para (algunos de) los propósitos para los que en la actualidad sirven las primeras: la suerte de relaciones que valen, p. ej., entre «cantidad de fluido calórico» y «energía cinética media de las moléculas». Hay un sentido obvio de «misma» en el que lo que se solía llamar «una cantidad de fluido calórico» es *la misma cosa* que lo que ahora se llama una cierta energía cinética media de las moléculas, pero no hay razón para pensar que todos los rasgos que se predican verdaderamente de lo uno pueden predicarse sensatamente de lo otro. (Rorty, 1965/1971, p. 176.)

Rorty intenta también diagnosticar por qué la gente se resiste comúnmente a aceptar los intentos de desembarazarse del vocabulario mentalista. Lo atribuye a lo poco práctico que resulta abandonar las antiguas expresiones idiomáticas en favor de un nuevo vocabulario científico[2]. Gran número de críticos está, sin embargo, en desacuerdo [139-140] con esta réplica. Cornman (1968) y Berastein (1968/1971) han defendido que, dado que el habla de sensaciones se usa en informes observacionales, el lenguaje que la reemplaza asumirá inevitablemente esta misma función de modo que no se eliminaría nada de modo efectivo. El nuevo discurso seleccionaría los mismos fenómenos mentalistas; simplemente emplearía palabras nuevas. Rorty ha rechazado esta afirmación. Mantiene que el contenido de aquello de lo que informamos es efectivamente una función de nuestro lenguaje y, de este modo, cambiará si cambiamos a un nuevo lenguaje: «si adoptamos el hábito de usar términos neurológicos en lugar de "intenso", "agudo" y "vibrante", entonces nuestra experiencia lo sería de cosas que tienen esas propiedades neurológicas y no de algo, p. e j ., intenso» (Rorty, 1970/1971, p. 228).

Más recientemente, Rorty (1979) ha intentado diferenciar su posición de la de Feyerabend centrándose en cómo tenemos conocimiento acerca de estados mentales, no en qué son. Considera como su objetivo primario la afirmación de que los fenómenos mentales son fenómenos a los que tenemos un acceso privilegiado. Mantiene que es esta idea del acceso privilegiado a nuestras mentes la que hace que la gente piense que hay una naturaleza esencial de los seres humanos. Rorty ha negado que tengamos tal

acceso privilegiado a lo que es humano. El lenguaje que usamos para describir nuestros estados mentales incorpora nuestras teorías sobre lo que es ser humano, y esas teorías representan decisiones basadas culturalmente. Las diferentes culturas pueden tomar decisiones diferentes respecto de lo que es una persona y las codificarán en su lenguaje. Ni la filosofía ni la ciencia pueden responder a la cuestión de lo que es ser una persona y decidir así qué lenguaje deberíamos usar. Una tarea de la filosofía es, de acuerdo con Rorty, exponer el hecho de que nuestras expresiones idiomáticas mentalistas codifican las decisiones que tomamos en nuestra cultura y no describen directamente la realidad de la vida mental.

El materialismo eliminativo jamás ha sido una posición altamente popular, pero aún tiene preeminentes defensores. Stephen Stich (1983) interpretó su teoría sintáctica de la mente (ver capítulo 4) como una posición eliminativista en la medida en que propone desarrollar la psicología científica sin apoyarse en modo alguno en la psicología popular intencional. Patricia y Paul Churchland, al avanzar [140-141] sus afirmaciones a favor de la neurociencia como nuestra mejor esperanza de desarrollar una ciencia viable de la mente, hacen afirmaciones que recuerdan, a menudo, las de Feyerabend y las de Rorty. Ellos han mantenido que, al continuar caracterizando los eventos mentales en términos de actitudes preposicionales, podemos estar estorbando nuestros esfuerzos de entender realmente los estados mentales. Mediante la investigación sobre cómo funciona el cerebro, afirman ellos, podemos aprender maneras mejores de describir nuestros estados mentales. En particular, Paul Churchland ha argumentado que por medio de la comprensión de los neuroprocesos que ocurren en el cerebro podemos enriquecer nuestra vida mental distinguiendo, por ejemplo, sonidos musicales que ahora confundimos. (Ver P.M. Churchland, 1981a, 1985, 1986; P.S. Churchland, 1980b, 1983, 1986; Churchland y Churchland, 1981. Discuto los puntos de vista de Churchland más completamente en Bechtel, en prensa b)[3].

El Materialismo Eliminativo, en la medida en que recomienda reemplazar las explicaciones mentalistas por otras de la neurociencia, tiene implicaciones negativas para gran parte del trabajo en ciencia cognitiva. Gran parte de la teorización en ciencia cognitiva emplea una perspectiva claramente mentalista (Palmer y Kimchi, 1986), que el Materialismo Eliminativo mantiene que es probablemente errónea. Si el Materialismo Eliminativo es correcto, deberíamos abandonar las investigaciones cognitivas y volver a dirigir los recursos a la neurociencia que tiene la mejor esperanza de explicar cómo opera la mente/cerebro.

La razón básica por la que el Eliminativismo no ha logrado una aceptación más amplia consiste en que los argumentos mentalistas desempeñan tal papel central en nuestro pensamiento ordinario sobre nosotros mismos, así como en las teorías de las ciencias sociales, que parece imposible que podamos pasar sin ellos. Kim (1985), por ejemplo, ha señalado algunos de los modos críticos en los que empleamos esta perspectiva mentalista:

> El esquema psicológico intencional —esto es, la armazón de creencia, deseo y voluntad— es aquel en el que deliberamos sobre fines y medios, y valoramos la racionalidad de las acciones y decisiones. Es la armazón que hace posibles nuestras actividades normativas y evaluativas. Ninguna armazón puramente descriptiva como la de la neurofisiología o la de la física, no importa cuán comprensiva y poderosamente predictiva sea, puede reemplazarla. En la medida en que podamos pensar sobre nosotros mismos [141-142] como agentes reflexivos capaces de evaluación y deliberación —esto es, en la medida en que nos consideramos a nosotros mismos como agentes

capaces de actuar de acuerdo con una norma—, no seremos capaces de prescindir de la armazón intencional de creencias, deseos y voliciones (p. 386).

Los defensores del Materialismo Eliminativo mantienen que tales afirmaciones a favor de nuestras expresiones idiomáticas mentalistas son simplemente conjeturas sobre qué dirección tomarán la ciencia y la sociedad. Lo que Kim hace claro, sin embargo, es que, al ofrecer un reemplazo para nuestra armazón mentalista. el eliminativista tiene que mostrar no sólo cómo podemos hacer psicología sin mentalismo, sino también cómo pueden funcionar sin él las ciencias sociales, y cómo los humanos pueden conducir sus vidas y determinar sus cursos de acción sin él. Aunque es posible un escenario en el que abandonemos nuestra concepción mentalista básica de los seres humanos y adoptemos la armazón conceptual de la neurociencia, nos parece algo profundamente implausible. (Para una discusión adicional, ver McCauley, 1986.)

Hay, además, algo problemático respecto del modo en que el Eliminativismo interpreta el problema. El eliminativista hace del asunto una cuestión de esto o lo otro: o mantenemos nuestra perspectiva mentalista o adoptamos la de la neurociencia, pero no ambas. Esto, sin embargo, puede ser confundir los problemas. Puede ser que las explicaciones de la neurociencia, e incluso el lenguaje de la neurociencia, enfoquen a un nivel diferente el discurso psicológico del sentido común. Considérese de nuevo la distinción de Dennett (discutida en el capítulo 4) entre la psicología intencional y la psicología del nivel del diseño y del nivel físico. Siguiendo a Dennett he argumentado que la psicología intencional desempeñaba un papel diferente que la psicología del diseño-postura. Aunque la última buscaba desarrollar modelos de cognición de procesamiento interno, la psicología intencional figuraba al explicar cómo un individuo se las había con su entorno, incluyendo otros agentes cognitivos. Gran parte de lo mismo puede aplicarse a la controversia sobre el Eliminativismo. Puede ser que podamos, a la vez, preservar el mentalismo y desarrollar una perspectiva propia de la neurociencia incluso si ambas cosas no logran engranar perfectamente. Las dos perspectivas servirán para propósitos diferentes. La posición final considerada en este capítulo, la Teoría de la Identidad como Instancia, intenta mostrar cómo pueden aceptarse las dos perspectivas aunque difieran entre sí. [142-143]

6.4 LAS TEORÍAS DE LA IDENTIDAD COMO INSTANCIA

Al igual que los materialistas eliminativos, los Teóricos de la Identidad como Instancia son escépticos respecto de la afirmación de la Teoría de la Identidad como Tipo de que la investigación apoyará una correlación entre tipos de fenómenos descritos mentalmente y tipos caracterizados físicamente, pero extraen una inferencia distinta de la que extraen los eliminativistas. Más bien que repudiar el discurso mental, los Teóricos de la Identidad como Instancia sancionan su uso continuo apoyando una versión débil de la Teoría de la Identidad. Mantienen que toda instancia de un evento mental es una instancia de un evento neural, pero no exigen que los tipos de eventos mentales se hagan equivaler con tipos de eventos neurales. Así pues, la Teoría de la Identidad como Instancia mantiene que: *a*) cada vez que estoy en un estado mental particular, ese estado mental es idéntico a un estado cerebral, pero *b*) en otras ocasiones, cuando estoy en el mismo estado mental, puedo estar en un estado cerebral diferente.

La posición de Donald Davidson, el Monismo Anómalo, ha sido una de las versiones más controvertidas de la Teoría de la Identidad como Instancia. La posición mantiene que el mismo evento puede ser a la vez mental y físico (de aquí el monismo), pero que no hay leyes que pongan en relación la descripción mental con la descripción física (de ahí que sea anómalo). Davidson (1970/1980; ver también 1973,1975) avanzó el Monismo Anómalo como un medio de reconciliar las tres tesis siguientes, todas las cuales eran consideradas por él como de aceptación obligada, pero que parecían inconsecuentes:

1. *El Principio de Interacción Causal*, que asevera que «al menos algunos eventos mentales interactúan causalmente con eventos físicos».
2. *El Principio del Carácter Nomológico de la Causalidad*, que enuncia que «donde hay causalidad, debe haber una ley: los eventos relacionados como causa y efecto caen bajo una ley estrictamente determinista».
3. *El Anomalismo de lo Mental*, que afirma que «no hay leyes estrictamente deterministas sobre cuyas bases puedan predecirse y explicarse eventos mentales» (Davidson, 1970/1980, pp. 80-81).

El monismo de Davidson mantiene que las actividades mentales son, cada una de ellas, idénticas con alguna actividad física (generalmente actividades neurológicas). Ésta es la afirmación de identidad crítica [143-144] que permite a Davidson satisfacer las dos primeras tesis. Puesto que todos los eventos mentales son eventos físicos, pueden interactuar causalmente con otros eventos físicos, y esas interacciones pueden caracterizarse mediante leyes físicas deterministas. La afirmación de que no hay leyes que pongan en relación descripciones mentales de eventos con sus descripciones físicas tiene la consecuencia de que no podemos inferir descripciones mentales a partir de sus descripciones físicas.

La resolución de Davidson de la supuesta incompatibilidad entre las tres tesis ha inspirado un gran número de críticas. Algunos críticos han objetado que Davidson no puede defender su afirmación de monismo puesto que no podemos establecer la identidad de lo mental y de lo físico sin ser capaces de relacionar tipos. Sin embargo, Davidson no está interesado en argumentar a favor de la identidad; él la postula simplemente como necesaria si hemos de acomodar las tesis 1 y 2. Su preocupación es, más bien, argumentar a favor de la carencia de leyes que relacionan las descripciones mentales y las físicas.

Al argumentar a favor del anomalismo, Davidson no niega que podríamos desarrollar generalizaciones que liguen los eventos descritos mentalmente con eventos descritos físicamente. Él afirma simplemente (Davidson, 1970/1980) que esas generalizaciones no tendrán el carácter de una ley:

> La tesis es que lo mental es nomológicamenie irreductible: puede haber enunciados generales verdaderos que ponen en relación lo mental y lo físico, enunciados que tienen la forma lógica de una ley: pero no son *legaliformes* (en un sentido fuerte que se describirá). Si por una casualidad absurdamente remota tropezásemos con una generalización psicológica verdadera de carácter no estocástico, no tendríamos razón alguna para creerla más que aproximadamente verdadera (p. 90).

La razón por la que el enunciado no es legaliforme consiste en que los predicados se extraerían de dos vocabularios diferentes que no pueden fusionarse en una ley. Davidson afirma que «los enunciados nomológicos ponen juntos predicados que sabemos que están hechos uno para el otro» (p. 93). Esto sólo

ocurre cuando se extraen «de una teoría con elementos constitutivos fuertes» (p. 94). Esta afirmación no establece por sí misma el anomalismo de lo mental, pues podemos pensar que hay elementos constitutivos fuertes que ligan predicados mentales y físicos o que podrían desarrollarse. La afirmación esencial en el argumento de Davidson es que tales conexiones son imposibles. Él mantiene que los principios divergentes que gobiernan nuestro uso del vocabulario físico y mental son tales [144-145] que no podríamos integrarlos en una teoría. Nuestro sistema de atribuciones mentales está gobernado por el principio de racionalidad, esto es: adscribimos creencias y deseos de tal manera que hacemos que las otras personas aparezcan como racionales. Para hacer esto tenemos que estar libres continuamente para reevaluar nuestras atribuciones de predicados mentales, y de este modo no podemos vincularlos fuertemente a propiedades físicas[4].

El argumento de Davidson en contra de la posibilidad de desarrollar principios constitutivos que liguen los vocabularios psicológico y psíquico parece colocar exigencias sobre tales principios que no aceptaríamos en otras áreas. En contextos donde los científicos han intentado unir los vocabularios de dos dominios diferentes (p. ej., el término *gene* de la genética y el término *cromosoma* de la citología), han reconocido que sus propuestas eran falibles y podrían tener que ser revisadas en la medida en que estuviera disponible nueva evidencia. Puede haber casos donde queramos que las teorías de una disciplina respondan sólo a las exigencias de esa disciplina sin estar constreñidas por las exigencias de otras disciplinas (ver Abrahamsen, 1987; McCauley, 1987b). Pero hay otras ocasiones en las que las constricciones impuestas por el resto, consecuentes con los compromisos teóricos de otras disciplinas, pueden ser una ventaja. Tales restricciones pueden ayudar a mostrar cuál de las dos teorías en competición dentro de una disciplina es más probable que sea verdadera. Además, tales constricciones pueden forzarnos a modificar los compromisos teóricos en una de las disciplinas. De hecho, éste es uno de los productos beneficiosos de los límites disciplinares cruzados y del trabajo de consulta en otra disciplina (ver Bechtel, en prensa b, capítulo 6; también McCauley, 1986). La prescripción de Davidson elimina tal beneficio de la teorización psicológica.

La razón de la fuerte oposición de Davidson a los principios que hacen de puente entre la psicología y la neurociencia reside en el hecho de que está comprometido con el principio de racionalidad como el único fundamento de nuestros intentos de interpretar a los agentes en términos psicológicos (ver Davidson, 1973). Detrás de la [145-146] posición de Davidson está una particular concepción de lo que incluye el discurso psicológico, una concepción que descarta el *status* de la psicología como una ciencia. Los principios de la psicología no son las bases para predecir o explicar la conducta (que requeriría leyes), sino para desarrollar explicaciones racionales de la conducta por medio de la interpretación de los agentes en términos de conjuntos coherentes de creencias y deseos. (Ver Lycan, 1981b, para una discusión crítica.)

La afirmación de Davidson de que la racionalidad proporciona el único criterio para juzgar las explicaciones psicológicas parece no sólo innecesaria sino también errónea. Podemos emplear la racionalidad como un criterio al desarrollar las explicaciones psicológicas sin exigir que sea el criterio absoluto. Reconocemos que tanto nosotros como otras personas somos algunas veces irracionales, pero esto no socava nuestra capacidad de desarrollar explicaciones que interpreten nuestra conducta como generalmente racional. Una estrategia importante en la ciencia es intentar identificar entidades de maneras múltiples de modo que los juicios basados en un modo de identificar las entidades puedan

comprobarse usando otros modos. Cuando esto es posible, los principios que emergen son más robustos y, por tanto, más creíbles (Campbell, 1966; Cook y Campbell, 1979; Wimsatt, 1981). La confianza de Davidson en la racionalidad sola como base para fijar interpretaciones mentales cierra de antemano esta opción. Sin embargo, si rechazamos el basarnos en la racionalidad sola, entonces socavamos también el apoyo a favor del monismo anómalo.

La versión de Davidson de la Teoría de la Identidad como Instancia deja lugar para las teorías cognitivas, pero con el coste de convertir en no científicas las explicaciones cognitivas. Sin embargo, otros filósofos que han ofrecido argumentos diferentes para favorecer a la Teoría de la Identidad como Instancia sobre la Teoría de la Identidad como Tipo presentan una versión de la Teoría de la Identidad como Instancia que es bastante más afín a la ciencia cognitiva. Fodor (1974) y Putnam (1975b) han ofrecido razones para pensar que las relaciones entre tipos mentales y tipos físicos son tales que el mismo evento mental puede realizarse, bajo circunstancias diferentes, en eventos físicos completamente diferentes. Fodor ha apelado al hecho de que clasificamos cosas diferentemente para propósitos diferentes. Por ejemplo, podemos clasificar los objetos por el color o por la forma, y no hay razón para pensar que las dos clasificaciones se correspondan entre sí. Similarmente, las clasificaciones útiles en psicología pueden ser completamente distintas de aquellas que son útiles para la neurociencia. Por ejemplo, puede ser útil en psicología social clasificar actividades de hacer promesas, que es una actividad que [146-147] puede realizarse mediante muchas actividades físicas diferentes que no es probable que formen un tipo físico[5].

Putnam ha ofrecido un argumento relacionado a favor de la afirmación de que un tipo mental dado de un evento mental puede realizarse por medio de diferentes eventos físicos. Apela al hecho de que, aunque hay diferencias modestas a lo largo del tiempo entre la constitución de nuestros cerebros y entre los cerebros de personas diferentes, adscribimos los mismos estados psicológicos a ellas. Además, los psicólogos comparativos están completamente preparados para adscribir los mismos estados psicológicos a miembros de especies diferentes cuyos cerebros son incluso más diferentes entre sí, y podemos imaginarnos adscribiendo los mismos estados a extraños con cerebros totalmente diferentes. Putnam (1978, 1983) ha contemplado también la posibilidad de que los mismos estados neurológicos puedan subyacer a propiedades psicológicas diferentes. El ha mantenido que la interpretación psicológica depende de consideraciones externas al sistema de modo que el mismo sistema en entornos diferentes será interpretado diferentemente. Aunque el enfoque de Putnam, que hace que las adscripciones psicológicas dependan de circunstancias del entorno, es altamente controvertido, pone a la vista uno de los factores centrales que ha motivado el desarrollo de la Teoría de la Identidad como Tipo: el hecho de que, aunque la psicología y la neurociencia puedan ambas caracterizar los mismos estados, pueden tener criterios diferentes para agruparlos en clases.

La Teoría de la Identidad como Instancia, tal como ha sido desarrollada por Fodor y Putnam, proporciona una explicación de la relación entre la mente y el cerebro que se compadece más que otras versiones del materialismo con las preocupaciones de la ciencia cognitiva, puesto que da cabida a un dominio autónomo para el teorizar cognitivo. Esta autonomía, sin embargo, puede también ser peligrosa si entraña, como algunos Teóricos de la Identidad como Instancia mantienen que entraña, que las teorías cognitivas son [147-148] inconmensurables con las teorías de la neurociencia, de modo que la ciencia cognitiva no puede aprender de la neurociencia ni servirle de guía. Una estrategia para relacionar la

ciencia cognitiva y la neurociencia que permite además alguna autonomía para la ciencia cognitiva se discute en Bechtel (en prensa b, capítulo 2).

6.5 RESUMEN DE LOS PUNTOS DE VISTA SOBRE EL PROBLEMA MENTE-CUERPO

En este capítulo y en el 5 he pasado revista a las posiciones filosóficas más importantes sobre la relación de la mente y el cuerpo y he explorado su significación para la ciencia cognitiva. El dualismo de objetos trata a las mentes como géneros de objetos radicalmente diferentes de los cuerpos físicos como el cerebro, y de esta manera coloca el estudio de la actividad mental fuera de los límites de la ciencia física. Al contrario, las otras posiciones colocan todas ellas a los fenómenos mentales dentro del dominio de la ciencia física, pero difieren en el modo como lo hacen. El conductismo filosófico argumenta que el discurso mental debería interpretarse como refiriéndose a la conducta o disposiciones a comportarse, y no a eventos internos del cerebro. Al negar el procesamiento interno, rechaza los géneros de explicación avanzados en la ciencia cognitiva contemporánea.

Las versiones del materialismo consideradas en este capítulo reconocen todas ellas alguna forma de procesamiento interno y a ese respecto son más consistentes con la investigación en la ciencia cognitiva. La Teoría de la Identidad como Tipo hace equivaler, sin embargo, los eventos mentales con eventos físicos que ocurren dentro del cerebro. Aunque la identificación de estados mentales con estados físicos asegura la realidad de los estados mentales que podría estudiar la ciencia cognitiva, la Teoría de la Identidad como Tipo entrañaría también que los procesos internos empleados en las explicaciones cognitivas serían isomórficos a los empleados en la neurociencia. Esto da primacía a la neurociencia sobre la ciencia cognitiva. El Materialismo Eliminativo se concentra de modo similar en los procesos neurológicos que ocurren en el cerebro, pero mantiene que, puesto que esas explicaciones neurales son inconsecuentes con las explicaciones cognitivas, deberíamos renunciar a las explicaciones cognitivas en favor de las neurales. Así pues, el Materialismo Elimininativo defendería abandonar la ciencia cognitiva en favor de la neurociencia.

La Teoría de la Identidad como Instancia afirma que puede haber una alternativa, explicaciones incompatibles de las actividades internas [148-149] de los sistemas cognitivos: una neural y otra cognitiva. Así pues, de todas las explicaciones filosóficas sobre el problema mente-cuerpo, la Teoría de la Identidad como Instancia es la más compatible con los programas de la ciencia cognitiva. La Teoría de la Identidad como Instancia plantea la cuestión de cómo han de categorizarse los eventos mentales si esta categorización ha de ser diferente de la que se aplica a los eventos cerebrales. Los defensores de la Teoría de la Identidad como Instancia han propuesto que los eventos mentales se defiendan funcionalmente. El programa filosófico conocido como *funcionalismo* ha intentado explicar lo que incluye esto. En consecuencia, el próximo capítulo está dedicado a examinar más estrechamente el funcionalismo. [149-150]

7. FUNCIONALISMO

7.1 INTRODUCCIÓN

El Funcionalismo representa un intento filosófico de explicar una parte crítica del programa de investigación de la ciencia cognitiva: el modo en que se reconocen y clasifican los eventos mentales. El Funcionalismo mantiene que los eventos mentales se clasifican en términos de sus papeles causales. Así pues, un evento mental se describiría en términos de su papel en el sistema mental, lo mismo que una palanca se caracteriza en términos de su papel causal consistente en controlar la apertura y el cierre de las válvulas del motor de un coche. Un aspecto importante de esta posición es que los eventos mentales pueden reconocerse y clasificarse independientemente de su constitución física. Por esta razón, el Funcionalismo se considera a menudo como incompatible con la Teoría de la Identidad como Tipo[1]. La posición sobre el problema mente-cuerpo con que más a menudo se empareja [150-151] el Funcionalismo es la de la Teoría de la Identidad como Instancia, que, del mismo modo, disocia las descripciones de los eventos mentales de aquellas que se aplican a los eventos físicos.

Usar el término *Funcionalismo* para este modo de clasificar eventos mentales tiende a causar confusión entre los científicos sociales y los de la conducta. En psicología, por ejemplo, el término se aplicó al programa de investigación desarrollado a finales del siglo pasado y principios de éste, especialmente en la Universidad de Chicago mediante el trabajo de Dewey y Angelí. Era un asunto clave de este enfoque una perspectiva evolucionista que llevó a los psicólogos a prestar atención al uso que un organismo hacía de sus capacidades cognitivas. Esta orientación evolucionista ha sido manifiesta en muchos de los enfoques de la psicología en el siglo XX, incluido el conductismo. La perspectiva evolucionista del conductismo psicológico no ha desempeñado un papel importante en el programa filosófico que subyace al mismo nombre. Sin embargo, en la última parte de este capítulo bosquejo una versión del Funcionalismo filosófico que introduce una perspectiva evolucionista.

Hay en la actualidad una variedad grande de versiones diferentes del Funcionalismo filosófico. Paso revista a cuatro de ellas en la primera sección. Aunque de una forma u otra el Funcionalismo ha atraído a un amplio espectro de adeptos, ha levantado también gran número de críticas. Así pues, en la segunda sección presento algunas de las principales objeciones al Funcionalismo y las respuestas que los funcionalistas han ofrecido. En la última sección desarrollo la versión alternativa mencionada previamente.

7.2 VARIEDADES DEL FUNCIONALISMO FILOSÓFICO

El hecho de que hay diversas variedades del Funcionalismo no es algo que se reconoce siempre, y la gente tiende a confundir las diferentes versiones. Esta situación puede ser particularmente confusa, puesto que los proponentes de una versión critican a menudo otras versiones (y algunas veces presentan sus críticas como críticas al Funcionalismo de manera general). Aunque todas las versiones del

Funcionalismo están de acuerdo en que los estados mentales han de identificarse primariamente en términos de sus interacciones mutuas, difieren principalmente sobre cómo han de especificarse esas interacciones. Comienzo con un punto de vista que identifica esas interacciones en términos de nuestro discurso mental ordinario y a continuación paso a puntos de vista que extraen su inspiración de las preocupaciones de la investigación contemporánea en ciencia cognitiva. [151-152]

7.2.1 FUNCIONALISMO DE LA PSICOLOGÍA POPULAR

El Funcionalismo de la Psicología Popular interpreta la armazón conceptual que se supone en el discurso de actitudes preposicionales (ver capítulo 3) como algo que incorpora una teoría sobre los factores causales que gobiernan la conducta humana. (Esta teoría se denomina *teoría popular* [*folk theory*] puesto que se supone que refleja el conocimiento común, no el conocimiento científico.) David Lewis (1972/1980) sugirió que los términos mentales como *desear* y *creer* se definen en términos de esta teoría. Para mostrar que hay realmente una teoría que subyace a las actitudes preposicionales, es necesario codificarla. Lewis propuso que esto podría hacerse articulando cierto número de perogrulladas de la psicología popular capturadas en el discurso de las actitudes proposicionales:

> Piénsese en la psicología del sentido común como una teoría científica introductora de términos, aunque inventada mucho antes de que hubiese una institución tal como la ciencia profesional. Colecciónense todas las perogrulladas en que uno pueda pensar respecto de las relaciones causales de los estados mentales, estímulos sensoriales y respuestas motrices. Quizás podamos pensar en ellas como algo que tiene la forma:
>
> Cuando alguien está en tal-y-tal combinación de estados mentales y recibe estimulación sensorial de tal-y-tal clase, él tiende con tal-y-tal probabilidad a ser causado mediante ello a pasar a tales-y-tales estados mentales y a producir tales-y-tales procesos motrices (p. 212).

Lewis contempló esta teoría como algo que determina el significado de nuestros términos mentales de la misma manera que otras teorías en otras disciplinas determinan los significados de sus términos componentes. Por ejemplo, en la mecánica newtoniana, los significados de términos como *masa* y *fuerza* se especifican en términos de leyes como «fuerza = masa x aceleración»[2].

Un problema del enfoque de Lewis es que, si su teoría fuese errónea, todo nuestro discurso sobre los eventos mentales se convertiría en vacío y sin referencia. (Ver Wilkes, 1981, que ha desarrollado ésta y otras críticas.) Armstrong (1968,1984) desarrolló una variante de este enfoque que evita apelar a cualquier teoría implícita. En lugar [152-153] de ello apela a un análisis de nuestro vocabulario mentalista ordinario para definir los términos mentales. Él mantiene que los significados de los diversos términos mentales afirman ciertas relaciones causales del mismo modo que términos como «elástico» o «frágil» especifican contingencias físicas causales. Parte de lo que queremos decir cuando adscribimos a una persona la creencia general de que, por ejemplo, «todos los F son G» es la expectativa de que, si la persona aprende que *a* es F, esto generaría causalmente la creencia de que *a* es G. Estas relaciones

causales definen lo que es para Armstrong tener un estado mental.

Una de las principales metas de esta forma de Funcionalismo es mostrar cómo entendemos el significado de los términos mentales ordinarios sin apelar al conductismo filosófico y sin conocer la naturaleza de los estados cerebrales subyacentes. Estos términos especifican un nexo de agentes causales que invocamos para explicar la conducta de agentes cognitivos[3]. El Funcionalismo de la Psicología Popular parece realizar esta tarea con éxito. La cuestión más profunda es, sin embargo, si esos análisis tendrán alguna utilidad al desarrollar explicaciones científicas de cómo operan los sistemas cognitivos. Algunos filósofos, como Fodor, consideran la psicología popular como un punto de partida para desarrollar tales teorías científicas. Pero, para desarrollar esos análisis como empeños científicos, es necesario ir más allá de los análisis del vocabulario psicológico ordinario. Necesitamos desarrollar nuevas perspectivas teóricas que puedan ser comprobadas empíricamente. Al desarrollar tales teorías científicas, los análisis funcionales pueden desempeñar un papel diferente, sugiriendo cómo han de estructurarse tales teorías. Dentro de ese espíritu se desarrollaron las siguientes tres versiones del Funcionalismo. [153-154]

7.2.2 FUNCIONALISMO DE TABLA DE MÁQUINA

El Funcionalismo de Tabla de Máquina es una de las primeras versiones del Funcionalismo desarrollada primariamente por Putnam (1960). Una Máquina de Turing (ver Turing, 1937) es un dispositivo simple que consta de:

1. una cinta de longitud potencialmente infinita que contiene una secuencia lineal de cuadrados, en cada uno de los cuales se puede escribir un conjunto finito de símbolos;
2. una unidad de ejecución, que puede estar en uno de un número finito de estados internos, y
3. un indicador que señala uno de los cuadrados de la cinta.

Las actividades de la unidad de ejecución están dirigidas por un conjunto finito de reglas condicionales que especifican que ha de realizarse una acción, dado el símbolo particular que aparece en el cuadrado indicado y el estado interno de la unidad de ejecución. La acción consiste en escribir el mismo símbolo o uno diferente en el cuadrado y mantener o cambiar el estado interno de la unidad de ejecución (ver figura 7.1). Si la máquina no tiene instrucciones para su estado presente y para el número que está sobre la cinta, entonces se para. La capacidad de operación total de una Máquina de Turing particular puede resumirse en una Tabla de Máquina que presenta las reglas convencionales que gobiernan la conducta del sistema. A una Máquina de Turing se le da un problema especificando los símbolos iniciales que hay sobre la cinta, y los símbolos que quedan sobre la cinta cuando la máquina se detiene (si es que lo hace) representan la solución del mismo. Putnam se interesó inicialmente en Máquinas de Turing porque la relación del programa que gobierna la Máquina de Turing parecía estar en una relación con el dispositivo físico que era, con mucho, igual a aquella en la que está la mente con el cerebro. Pensó que apelando a esta analogía podría disipar gran parte de las preocupaciones sobre el *status* ontológico de la mente, puesto que parece que no hay razón alguna para ser dualista respecto de una Máquina de Turing y los casos parecen completamente comparables.

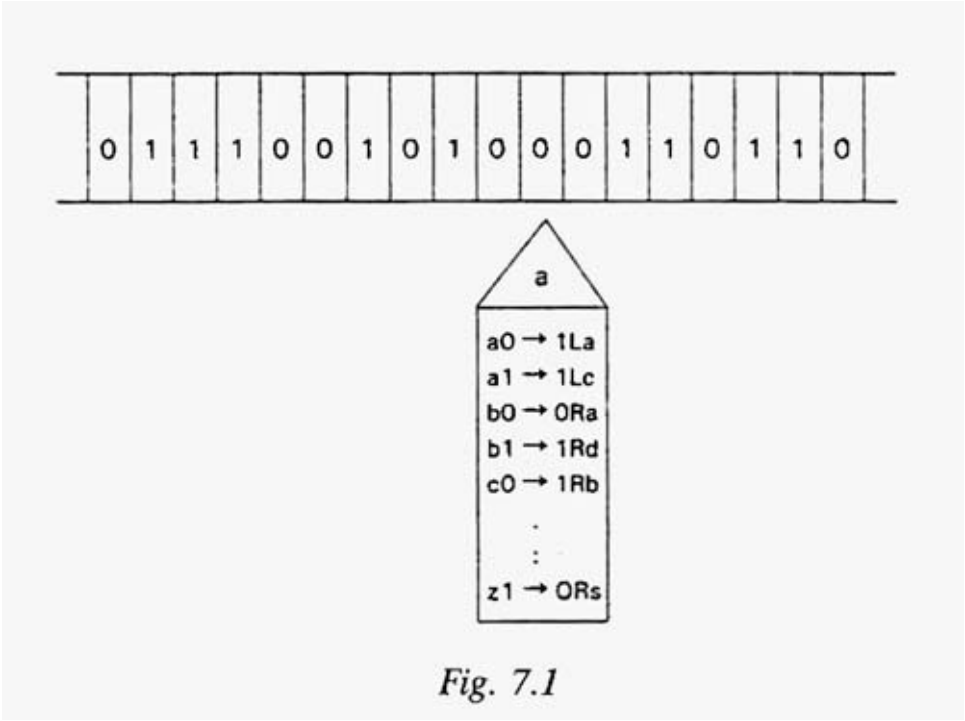


Fig. 7.1

TEXTO DE LA FIGURA 7.1. Una Máquina de Turing simple. En cada cuadrado de la cinta de longitud potencialmente infinita aparece o un *0* o un *1*. El indicador de la unidad de ejecución señala un cuadrado que contiene un *0*. La *a* de la parte triangular de la unidad de ejecución indica que la unidad de ejecución está en el estado *a*. Las reglas condicionales que gobiernan la actividad de la unidad de ejecución están enunciadas dentro de la parte en recuadro de la unidad de ejecución. La tetra y el número que está antes de la flecha especifican las condiciones bajo las que se aplica la regla (p. ej., cuando el ejecutor está en el estado *a* y el indicador está señalando *0*). La secuencia que hay después de la flecha indica qué número debe escribir la unidad de ejecución sobre el cuadrado al que está señalando, si se debe mover a la izquierda o a la derecha y en qué estado debe entrar entonces. La primera regla, que se aplica a la situación representada, dice a la cabeza que escriba un *1* sobre el cuadrado al que está señalando, que se mueva un cuadrado a la izquierda y que pase al (permanezca en el) estado *a*.

Las Máquinas de Turing han adquirido un interés adicional en las discusiones sobre el carácter de la mente como resultado del argumento, debido a Turing y Church, de que en la medida en que los procesos llevados a cabo por la mente son efectivamente computables, hay una Máquina de Turing que será conductistamente equivalente [154-155] a la mente[4]. Esto sugiere que podemos especificar las actividades de la mente en una tabla de Máquina de Turing y que podríamos identificar los estados mentales con estados o disyunción de estados de una máquina especificados por la tabla de máquina (de ahí el [155-156] nombre de «Funcionalismo de Tabla de Máquina»). Putnam (1967/ 1980) aplicó el análisis de tabla de máquina al dolor. Propuso que el dolor, más bien que ser un estado del cerebro, era un estado o conjunto de estados de un sistema que resulta de ciertas suertes de *inputs*, donde la conducta global del sistema se especifica por una tabla de máquina. Putnam distinguió también esta propuesta del conductismo filosófico afirmando que su análisis no exige una traducción del discurso de dolor en ningún discurso disposicional. Más bien, el dolor se hace equivaler a un estado del sistema que, de acuerdo con la tabla de máquina, produce causalmente otros estados dentro del sistema, así como *outputs* a partir del sistema.

El Funcionalismo de Tabla de Máquina ha levantado una gran variedad de objeciones incluso entre aquellos que se cuentan generalmente a sí mismos como funcionalistas. Block y Fodor (1972/1980) se han quejado de que hay un gran número de rasgos de los fenómenos psicológicos que no pueden manejarse satisfactoriamente por el Funcionalismo de Tabla de Máquina. Por ejemplo, no puede captar la importante distinción entre estados mentales que ocurren efectivamente (contemplar efectivamente la proposición de que, si hay nubes que amenazan lluvia y truenos, entonces la lluvia vendrá a continuación) y estados disposicionales (creer pero no contemplar activamente la proposición de que; si hay nubes que

amenazan lluvia y truenos, entonces la lluvia vendrá a continuación). Esto es así puesto que todos los estados indicados en la tabla de máquina son de un género. Una objeción adicional consiste en que una explicación de máquina de tabla individuará estados mentales demasiado finamente, puesto que distinguirá estados en dos autómatas si hay alguna diferencia o en las condiciones de *input* o en las condiciones de *output* para un estado particular sin importar lo trivial que sea. Una objeción adicional es que los estados de la tabla de máquina tienen que ser finitos, mientras que el número de estados psicológicos es potencialmente infinito.

Sin embargo, Block y Fodor sugieren también cómo vencer esos problemas del Funcionalismo de Tabla de Máquina. El Funcionalismo de Tabla de Máquina trata a un estado de máquina como algo comparable al estado psicológico total de una persona. La clave de su propuesta es identificar los estados mentales con estados computacionales dentro del sistema, donde los estados se definen en términos de géneros de operaciones realizadas. El resultado es que no identificaremos un estado mental con un estado de una máquina particular, sino con una operación que podría realizarse en una gran variedad de máquinas. Desarrollar el análisis de esta manera nos permite diferenciar entre procedimientos disponibles en la máquina (que podrían [156-157] compararse con estados disposicionales) y aquellos procedimientos que se realizan efectivamente. Además podemos comparar procedimientos de dos máquinas incluso cuando otros procesos de las dos máquinas difieren. Los dos procedimientos cuentan como el mismo si pueden substituirse uno por otro sin cambiar otras actividades dentro del sistema. Fodor y Block avanzan entonces el Funcionalismo Computacional como reemplazo del Funcionalismo de Tabla de Máquina.

7.2.3 FUNCIONALISMO COMPUTACIONAL o DE IA

Esta forma de Funcionalismo está estrechamente asociada a la Teoría Computacional de la Mente discutida en el capítulo 4. La mente se contempla aquí como algo que lleva a cabo operaciones sobre símbolos codificados dentro de ella. Haugeland (1981/1985) ha caracterizado el punto de vista resultante sobre la mente como un sistema formal automático interpretado. Un sistema formal es simplemente aquel en el que símbolos discretos se manipulan de acuerdo con un conjunto finito de reglas. Estas reglas establecen diferencias entre símbolos en virtud de rasgos formales tales que una regla específica manipulará dos símbolos formalmente equivalentes de la misma manera. Un sistema automático es aquel en el que las reglas que gobiernan la manipulación de símbolos se incorporan dentro del sistema y no tienen que proporcionarse de manera continuada por un agente externo. Finalmente, un sistema formal automático se interpreta cuando a sus símbolos se les proporciona una semántica, esto es: se considera que se refieren a cosas externas al sistema. Usando la expresión *sintaxis* para referirse a las propiedades formales y *semántica* para la interpretación, Dennett (1981a) caracterizó esta versión del Funcionalismo como aquella en la que la mente se contempla como un motor sintáctico que emula un motor semántico.

El Funcionalismo Computacional o de IA se compromete entonces a caracterizar las actividades mentales en términos de símbolos y reglas para manipular esos símbolos. Para emplear esta armazón a fin de comparar sistemas, especialmente para comparar computadores con humanos, tenemos que hacer precisas las nociones de símbolo y de regla de manera que podamos determinar cuándo dos sistemas están empleando el mismo conjunto de reglas o representaciones y cuándo están usando conjuntos

diferentes. La razón por la que esto es necesario puede reconocerse considerando una prueba propuesta para comparar humanos y máquinas: la Prueba de Turing (Turing, 1950/1964). Esta prueba aceptaría que un computador es [157-158] comparable en inteligencia a un ser humano si los seres humanos no pueden distinguir las realizaciones de un computador de las de un ser humano. La noción de Equivalencia de Turing desarrollada a partir de esta prueba hace equivaler dos sistemas que producen el mismo *output* a partir del mismo *input*. La Equivalencia de Turing no establece, sin embargo, que los dos sistemas funcionan de la misma manera, puesto que considera sólo los *outputs* de conducta y no si se emplean los mismos procedimientos internos (reglas y símbolos). Una gran variedad de sistemas de computador puede producir los mismos outputs globales usando secuencias de pasos muy diferentes. Esta capacidad para desarrollar diferentes sistemas de reglas para computar la misma función da lugar a una distinción que se ha hecho alguna vez entre dos enfoques de la *inteligencia artificial*. Un enfoque, que ha tomado el nombre genérico de inteligencia artificial, considera que su tarea es simplemente diseñar máquinas que puedan realizar funciones cognitivas sin preocuparse demasiado por la cuestión de si las realizan de una manera semejante en alguna medida a como los humanos las realizan. El otro, que ha adoptado el nombre de *simulación cognitiva*, considera como un objetivo principal el desarrollo de máquinas que realicen funciones cognitivas del mismo modo en que las realizan los humanos. Esta distinción es importante para el Funcionalismo Computacional. Si el Funcionalismo Computacional ha de ser una explicación de la cognición humana, entonces la meta debe ser una simulación cognitiva donde los programas de computador efectúen las mismas operaciones que los seres humanos.

«Efectuar la misma operación» se caracteriza en la jerga de computador como seguir el mismo *algoritmo*, donde un algoritmo especifica simplemente una secuencia de pasos, donde cada uno de los pasos constituye un procedimiento primitivo. Sin embargo, para comparar los algoritmos necesitamos una especificación de los procedimientos primitivos para los que se construyen los algoritmos. Éstos los proporciona, para los programadores de computador, el lenguaje en el que está escrito el programa. La mayor parte de los programas de computador están escritos en lenguajes de computador de alto nivel, cada una de cuyas instrucciones se traduce, mediante procedimientos conocidos como «compilación» e «interpretación», en un conjunto específico de operaciones de un lenguaje de nivel inferior. En última instancia las instrucciones tienen que traducirse a un código de máquina cuyos símbolos primitivos especifican directamente operaciones físicas dentro del computador. Aunque los programadores expertos pueden moverse libremente entre los lenguajes de nivel superior y los lenguajes de nivel inferior en los que el de nivel superior está siendo interpretado o compilado, entretejiendo varios niveles [158-159] dentro del mismo algoritmo, la mayor parte de los programadores permanece al mismo nivel. Para ellos el lenguaje de programación puede pensarse como la especificación de las operaciones primitivas disponibles en la máquina y, de este modo, se habla de él como definiendo una «máquina virtual».

Pylyshyn (1980,1984) ha apelado al concepto de máquina virtual al desarrollar una armazón para comparar programas. La máquina virtual proporciona la *arquitectura funcional* de la máquina. El ha propuesto que podemos comparar las operaciones de diferentes computadores o las de un computador con las de un ser humano en términos de lo que se especifica en la arquitectura funcional:

> puede pensarse en dos programas como estrechamente equivalentes o como realizaciones diferentes del mismo algoritmo o del mismo proceso cognitivo si se pueden representar por el

> mismo programa en alguna máquina virtual especificada teóricamente. Una manera simple de enunciar esto es decir que individuamos los procesos cognitivos en términos de sus expresiones en el lenguaje canónico de esta máquina virtual. La estructura formal de la máquina virtual —o lo que yo llamo su *arquitectura funcional*— representa entonces la definición teórica de, por ejemplo, el nivel correcto de especificidad (o nivel de agregación) en el que contemplamos los procesos mentales, el género de recursos funcionales de que el cerebro dispone, cómo están organizados la memoria y su acceso, qué secuencias se permiten, qué limitaciones existen sobre el paso de argumentos y sobre las capacidades de distintos amortiguadores, y así sucesivamente. (Pylyshyn, 1984, p. 92.)

En el caso de un computador, la arquitectura funcional no es absoluta. Podemos cambiar las capacidades primitivas de la máquina proporcionando un intérprete o un compilador para introducir un lenguaje de nivel superior, o entrando directamente en un lenguaje de nivel inferior. Pero, en el caso de los humanos, Pylyshyn ha argumentado que hay una arquitectura cognitiva básica que está disponible gracias a la constitución biológica del sistema nervioso. Para Pylyshyn, la tarea primaria de la ciencia cognitiva consiste en descubrir la estructura de esta arquitectura. Solamente cuando sepamos en qué consiste esta arquitectura seremos capaces de especificar cuáles son las operaciones primitivas básicas, y de este modo tendremos una base para comparar los procesos en el computador y en el ser humano.

Pylyshyn ha propuesto dos métodos para descubrir la arquitectura funcional de la mente humana. Uno incluye desarrollar simulaciones en las que los expedientes utilizados para diferentes tareas exigen (en tanto que medidos, p. ej., en tiempos de procesamiento) ser correlacionados con los que se encuentran cuando los seres humanos [159-160] llevan a cabo las mismas tareas. Aunque los tiempos absolutos diferirán entre humanos y máquinas, si las cantidades de tiempo requeridas para diferentes tareas exhiben la misma *ratio* en el humano que en el computador, entonces parece razonable suponer que se basan en operaciones básicas comparables. Esto, sin embargo, no revela directamente el nivel de arquitectura funcional, puesto que es posible que las operaciones sean interpretadas o compiladas en otras más básicas. Nos dice sólo que la comparación entre sistemas es apropiada en algún nivel. El otro enfoque de Pylyshyn consiste en identificar la arquitectura funcional con aquellas operaciones cuya realización no puede ser alterada por la información. Él habla de esas operaciones como «cognitivamente impenetrables». La idea es que, si la información puede alterar la realización, entonces la operación no está fijada sólo por la biología[5]. La dificultad es encontrar operaciones fijadas de esta manera, especialmente si el sistema biológico es un sistema adaptativo que puede modificarse como respuesta al procesamiento cognitivo.

Aunque quedan algunas dificultades a la hora de establecer cuándo una máquina y un humano procesan información de la misma manera, la estrategia del Funcionalismo Computacional o de la Inteligencia Artificial es clara. Una vez que identificamos un conjunto comparable de procedimientos básicos que pueden ejecutarse a la vez por mentes y por computadores, debemos intentar diseñar e implementar algoritmos en el computador que use la misma secuencia de operaciones que las seguidas por la mente humana. El Funcionalismo Computacional se compromete entonces con lo que Searle (1980/1981) denomina «IA fuerte». El computador no es simplemente un instrumento para ejecutar programas que caracteriza la conducta de la mente (como un computador podría ejecutar un programa

para determinar la trayectoria de un cometa sin pasar por los mismos procesos que pasa el cometa). Los computadores y la mente realizan procedimientos equivalentes al producir sus conductas, y esos procedimientos son lo que esta versión del Funcionalismo hace equivaler con procesos mentales. (Ver Anderson, Greeno, Kline y [160-161] Neves, 1981, para un intento de satisfacer las exigencias de la IA fuerte. Para una discusión filosófica y una evaluación de los intentos de modelar la actividad mental en inteligencia artificial, ver Boden,1977.)

Una objeción bastante común al Funcionalismo Computacional o de Inteligencia Artificial es que parece contemplar la mente como algo comparable a los computadores de von Neumann contemporáneos donde una instrucción tiene acceso de manera sucesiva a la información almacenada en la memoria y dirige las operaciones que han de realizarse para llevar a cabo la tarea de que se trate. Pero ahora resulta claro que el sistema nervioso humano realiza un gran número de operaciones al mismo tiempo y que puede no haber en él nada comparable a la función ejecutiva, tal como está incorporada de modo habitual en los computadores de von Neumann. Sin embargo, esta objeción malinterpreta de manera fundamental aquello con lo que está comprometido el Funcionalismo Computacional. No está comprometido con la afirmación de que las operaciones de la mente sean comparables a las de un computador de von Neumann. Una de las razones por las que Pylyshyn se preocupa por la arquitectura funcional de la mente es porque espera que, con toda probabilidad, sea diferente de la que se encuentra en los computadores contemporáneos. Él mantiene que tenemos que descubrir la arquitectura apropiada y desarrollar computadores usando esa arquitectura (bien como su arquitectura básica o como una arquitectura compilada o interpretada en otra) antes que la tarea de hacer modelos serios de la cognición humana pueda llevarse a cabo. Aquello con lo que está comprometido el punto de vista computacional es la afirmación de que la cognición es una actividad de procesamiento de símbolos. Sin embargo, este compromiso está siendo desafiado también por aquellos que desarrollan modelos no simbólicos en inteligencia artificial, tales como máquinas de procesamiento distribuido en paralelo (ver la discusión en el capítulo 4). Si se demostrase que esos desafíos basados en modos no simbólicos son correctos, entonces el Funcionalismo Computacional tendría graves dificultades.

7.2.4. FUNCIONALISMO HOMUNCULAR

El Funcionalismo Homuncular toma su nombre del punto de vista, a menudo ridiculizado, de que las funciones cognitivas son realizadas por un hombrecillo (*homunculus*) que tenemos dentro de la cabeza. El problema de este enfoque es que no explica de hecho nada sobre la cognición. Tenemos que explicar todavía cómo ese hombrecillo [161-162] realiza sus actividades cognitivas, y corremos el riesgo de un regreso al infinito en el que el hombrecillo necesita su propio *homunculus*, y así sucesivamente; El Funcionalismo Homuncular apoya la postulación de homúnculos dentro de una persona, pero evita la objeción del regreso postulando un gran número de ellos, cada uno de los cuales es más estúpido que el sistema global pero realiza una tarea necesaria para todo el sistema. Dennett caracteriza a los Funcionalistas Homunculares como aquellos que obtienen y a continuación devuelven préstamos de inteligencia. El préstamo se obtiene cuando caracterizamos al sistema y a sus homúnculos como inteligentes. Devolvemos este préstamo tomando nuevos préstamos, postulando un equipo interno de homúnculos dentro de cada homúnculo. Esto constituye un cierto progreso, puesto que cada nivel de

homúnculos requiere homúnculos de menor inteligencia. Repetimos el proceso hasta que alcanzamos homúnculos que requieren tan poca inteligencia, que podemos reemplazarlos por máquinas, y así se paga todo el préstamo[6].

Subyacente al Funcionalismo Homuncular está la concepción de la explicación científica que Cummins (1975,1983) ha denominado «explicación funcional»[7]. La meta de la explicación funcional es responder a la pregunta «¿En virtud de qué S tiene P?». Cummins ha sugerido que el modo apropiado de responder a tal cuestión es «construir un análisis de S que explique la posesión por parte de S de P apelando a las propiedades de los componentes de S y a su modo de organización» (Cummins, 1983, p. 15). Los componentes pueden identificarse de dos maneras: o físicamente o en términos de sus capacidades para interactuar con otros componentes. Lo último constituye [162-163] un análisis funcional que puede representarse por medio de una carta de flujo que muestra cómo la actividad global del sistema resulta a partir de la realización de operaciones dentro del sistema.

Hasta aquí mi caracterización del Funcionalismo Homuncular ha intentado distinguirlo del Funcionalismo Computacional, pero la idea de una carta de flujo sugiere un modo de relacionar los dos puntos de vista. Las actividades asignadas a los recuadros en una carta de flujo son tareas para las que los programadores intentarían escribir subrutinas. Dennett (1975/1978) mismo ha desarrollado esta comparación:

> El investigador de IA *empieza* con un problema caracterizado intencionalmente (p. ej., ¿cómo puedo hacer que un computador *entienda* preguntas en castellano?), lo desmenuza en subproblemas que se caracterizan también intencionalmente (p. ej., ¿cómo puedo hacer para que el computador *reconozca* preguntas, *distinga* sujetos de predicados, *ignore* configuraciones irrelevantes?) y a continuación vuelve a desmenuzarlos adicionalmente hasta que alcanza problemas o descripciones de tareas que son obviamente mecánicas (p. 80).

Aunque existe esta afinidad entre el Funcionalismo Homuncular y el Funcionalismo Computacional, el punto focal es diferente. La meta de la mayor parte de los investigadores de IA[8] es sintética: diseñar un programa para realizar la tarea global. La estructura jerárquica de un programa no es, en última instancia, crítica, y las operaciones de las subrutinas son de una pieza respecto de las operaciones del programa principal. Sin embargo, para el Funcionalista Homuncular la estructura jerárquica resulta más importante. El Funcionalista Homuncular trata los recuadros de las cartas de flujo como algo que caracteriza unidades modulares efectivas que efectúan sus propias actividades. Lo mismo que al sistema global se le atribuyen creencias y deseos, los funcionalistas homunculares como Dennett también atribuyen creencias y deseos a los homúnculos que constituyen el sistema. Las creencias y deseos de un homúnculo serán diferentes de las que se atribuyen ai sistema total: serán creencias y deseos sobre las tareas a realizar por el homúnculo (ver también Lycan, 1981a, 1981c).

Además de proponer un punto de vista jerárquico de la cognición, el Funcionalismo Homuncular no insiste en una única distinción entre función y estructura. Compara un sistema con un conjunto de cajas chinas unas dentro de otras. A medida que abrimos cada una [163-164] de ellas, entramos a un micronivel inferior. Ese proceso continúa de la misma manera hasta que estamos abajo, en el nivel neurofisiológico. Lycan (1981a) ha propuesto, de acuerdo con esto, que la Teoría de la Identidad se convierte en un caso

especial de Funcionalismo Homuncular:

> si aceptamos también mi afirmación de que las caracterizaciones homunculares y las caracterizaciones fisiológicas de los estados de personas reflejan meramente diferentes niveles de abstracción dentro de una jerarquía o continuo funcional circundante, entonces ya no podemos distinguir al funcionalista del teórico de la identidad de ninguna manera absoluta. «Neurona», por ejemplo, puede entenderse o como un término fisiológico (que denota un género de célula humana) o como un término (teleo-)funcional (que denota un relé de carga eléctrica); en *cualquiera de las dos* interpretaciones está por algo instanciable —si se quiere, por el papel desempeñado por un grupo de objetos más fundamentales—. Así, pues, *incluso el teórico de la identidad es funcionalista*: alguien que coloca a las entidades mentales en un nivel muy bajo de abstracción (p. 47).

7.3 OBJECIONES AL FUNCIONALISMO

En la sección anterior he descrito cierto número de versiones del Funcionalismo. A pesar de las diferencias entre ellas, comparten una suposición común: que lo que define los estados mentales es su capacidad para interactuar entre sí. Aunque esta perspectiva ha sido adoptada por muchos filósofos, ha sido objeto también de muchas críticas. En esta sección discuto varias objeciones que se han planteado contra el Funcionalismo y algunas de las réplicas que se han presentado a su favor.

7.3.1. OBJECIONES A LOS ANÁLISIS CAUSALES Y MECÁNICOS

Un tipo de objeción al Funcionalismo desafía el intento de caracterizar causalmente los estados mentales. Los conductistas filosóficos, como Malcolm, mantienen no sólo que la conducta y las disposiciones de conducta proporcionan el criterio para atribuir estados mentales a la gente, sino también que existe una conexión lógica entre estados mentales y conducta. Si una relación es lógica, no puede ser causal, puesto que las relaciones causales son contingentes y se descubren empíricamente mientras que las relaciones lógicas no son así. Respecto de cualquier relación causal, seria al menos concebible que la presunta causa no produjese el efecto, Pero esto no es posible cuando los eventos están relacionados lógicamente. Así [164-165] pues, Malcolm (1984) ha afirmado, por ejemplo, que un estado de pánico está relacionado lógicamente con lo que indujo el pánico. Por consiguiente, no podemos pensar que el pánico ocurra sin la circunstancia que lo ha inducido y, de este modo, la relación entre la circunstancia que lo ha inducido y el pánico no puede ser causal.

El Funcionalista, sin embargo, rechaza la afirmación de que la relación entre estados mentales y la conducta sea un asunto de lógica. Nosotros podríamos usar la conducta para identificar el estado mental de alguien, pero esa identificación es falible. Un estado mental estará conectado con una gran variedad de estados mentales diferentes en esta red causal. Esto abre la posibilidad de que seamos capaces de identificar el estado mental en una gran variedad de maneras diferentes. Cualquiera de esas maneras

puede considerarse como falible, revisable si otras maneras de identificar estados mentales nos llevan a una conclusión diferente. Cuando diversos indicadores diferentes señalan en la misma dirección, entonces nuestra evidencia a favor del estado mental es más fuerte y más fiable que cuando debemos fiarnos de sólo un indicador (para una discusión de la importancia de resultados fuertes en el desarrollo de la ciencia, ver Wimsatt, 1981). Para reconsiderar el ejemplo de Malcolm, si una gran variedad de criterios conductistas y psicológicos indican todos ellos que una persona está en un estado de pánico, entonces, incluso sin conocimiento del estado inductor del pánico, podemos decir que la persona es presa del pánico. Inversamente, incluso si sabemos que una persona ha experimentado una circunstancia que induce usualmente al pánico, si otros criterios no confirman la existencia de pánico, nosotros decidiremos que la circunstancia en cuestión no ha causado pánico.

El uso del computador para simular procesos mentales ha provocado una ráfaga de otras objeciones antimecanicistas. Se mantiene que contemplar los procesos cognitivos como procesos mecánicos parecidos a los de un computador es deshumanizador. Sin embargo, Boden (1981) ha argumentado que las simulaciones de computador son, con mucho, menos deshumanizadoras que los anteriores modelos mecánicos, puesto que postulan procesos internos. Esos procesos internos son análogos a los estados subjetivos postulados en los análisis más humanistas de la conducta humana. En tanto que poseer estados internos subjetivos es un aspecto importante de nuestro sentido de nosotros mismos como humanos, el modelo del computador y el análisis funcionalista que lo acompaña humanizan más que deshumanizan.

Un desafio bastante común a las simulaciones computacionales de la mente consiste en afirmar que los computadores pueden sólo comportarse en la medida en que están programados. Pero el pensamiento [165-166] humano, se asevera, no está constreñido de esa manera, puesto que es capaz de creatividad. Hay dos maneras de responder a esta objeción. Una es cuestionarse si los seres humanos no podrían estar programados en el sentido relevante. Incluso los empiristas más devotos admiten que los humanos vienen equipados para procesar información de ciertas maneras. Esos procedimientos para procesar —que son de nacimiento— podrían constituir un programa. Además, a los humanos se les enseña generalmente cómo llevar a cabo una gran variedad de actividades. Este proceso de instrucción podría contemplarse como algo comparable a la programación. La segunda consiste en desafiar la idea de que un programa constriñe del modo en que supone la objeción. Algunos programas están cerrados puesto que especifican todas las respuestas que el computador dará. Pero otros programas están abiertos puesto que se modifican dependiendo de los resultados de la ejecución del programa. Tales programas pueden estar estructurados de modo que generen, y seguidamente pongan a prueba, nuevas variaciones de sí mismos (variaciones que no estaban contempladas explícitamente por el programador). Parece al menos plausible que los computadores programados para generar nuevas estrategias y evaluarlas exhiban creatividad del mismo modo, en gran medida, que los humanos. El peso de la prueba, al menos, parecería estar sobre la persona que mantiene que los humanos son distintos: ella debería mostrar en qué consiste esta diferencia[9]. [166-167]

7.3.2. OBJECIONES A LAS EXPLICACIONES FORMALES DE LOS PROCESOS MENTALES

Searle y Dreyfus han sido dos de los críticos más prominentes del Funcionalismo, argumentando ambos en contra de la idea de que cognición consiste en procesamiento formal de símbolos solos. He

indicado ya la naturaleza de sus objeciones en el contexto de la evaluación de la Teoría Computacional de la Mente como una explicación de la intencionalidad. Puesto que Dreyfus (1979) ha discutido explícitamente los intentos particulares de la IA de explicar la cognición en términos de modelos formales, vale la pena considerar aquí su posición con más detalle. Dreyfus mantiene qué es bastante probable que sea imposible dar cuenta de la cognición humana en términos de representaciones y reglas formales para procesarlas. Ha argumentado a favor de esta afirmación examinando dos programas de investigación en inteligencia artificial: programas para tratar con mundos limitados (micromundos), especialmente diseñados, y programas que usan estructuras de conocimiento de nivel superior.

Dreyfus ha examinado el programa SHRDLU de Winograd (1972) como un ejemplo del proyecto del micromundo. Este programa estaba disenado para discurrir sobre un mundo hipotético de bloques. Aunque este programa era capaz de seguir la pista a los movimientos de bloques y responder correctamente a varias preguntas sobre los bloques, Dreyfus objetaba que este programa era sólo útil en el micromundo y que no podía generalizarse de manera que trate de un dominio más amplio. Como Haugeland (1985) ha observado, hay toda una hueste de cuestiones sobre bloques que SHRDLU no puede responder puesto que carece de los conceptos involucrados. El programa está equipado con un procedimiento para aprender a aplicar nuevos conceptos al mundo del bloque, pero necesita que se le enseñe cada uno de esos conceptos individualmente. Generalizar este programa para tratar con el mundo real exigiría introducir definiciones para aplicar cada concepto en cada uno de los dominios que en un número casi infinito existen. Dreyfus consideró esto como una evidencia de que el programa no seguía la dirección correcta. Nuestra capacidad para operar en el mundo no descansa en la combinación de elementos modularizados de información, cada uno de los cuales es aplicable a un micromundo específico. Dreyfus afirma que un mundo efectivo consta de «un cuerpo organizado de objetos, propósitos, destrezas y prácticas en términos de los cuales tienen significado las actividades humanas». «Aunque hay afirma él -un mundo infantil en el que, entre otras cosas, existen bloques, no hay tal cosa como un mundo de bloques» (Dreyfus, 1979, p. 163). [167-168]

Dreyfus contempla los afanes de Marvin Minsky, Roger Schank y otros de desarrollar estructuras de conocimiento que integren porciones individuales de información como mejoras respecto del programa de los micromundos. Los contornos de Minsky (1975/1981) constan de nudos donde se codifica información particular y de relaciones entre nudos. En un contorno para «boda», por ejemplo, algunos de los nudos contienen información que es siempre verdadera de las bodas (p. ej., incluyen dos partes que se casan), mientras que otros contienen información típicamente verdadera pero que puede ser alterada (p. ej., las dos partes tienen invitados). Incorporando valores por defecto modificarles para algunos nudos, tales estructuras son capaces de dirigir investigación y búsqueda activas para determinar lo apropiado de varios valores por defecto en contextos particulares.

A pesar de que representa una mejora, Dreyfus encuentra este enfoque inapropiado en última instancia, puesto que exige todavía control interno completo de la actividad del sistema. Dreyfus ha argumentado que en la actividad humana hay muchos factores que se mantienen externos, especialmente aquellos que controlan el modo en que el sistema hace frente a lo que le rodea y obtiene conocimiento de ello. La objeción de Dreyfus puede verse considerando un intento de desarrollar una estructura de conocimiento (un guión) para ir a un restaurante (Schank y Abelson, 1977). El guión especifica las actividades típicas incluidas en ir a un restaurante, pero permite que se den ciertas variaciones que

ocurrirán entre diferentes géneros de restaurantes (p. ej., restaurantes continentales y orientales). El guión intenta representar todos los aspectos relevantes de la experiencia del restaurante. Dreyfus afirma que esto es simplemente inútil puesto que una actividad humana como ir a un restaurante está regulada en gran medida por factores que no se representan internamente. Algunos de ellos pueden ser factores del medio externo a los que somos sensibles, mientras que otros son prácticas aprendidas (p. ej., seguir a la camarera hasta el lugar para sentarse). Puesto que el programa de Schank para responder cuestiones sobre ir a un restaurante no es sensible a tales influencias, el programa no sabe realmente sobre restaurantes. Para comprender la cognición la ciencia cognitiva, de acuerdo con Dreyfus, tiene que trabajar con sistemas *de carne y hueso*, no con sistemas formales abstractos.

La posición de Dreyfus (que él ha extraído de Heidegger) de que alguna información permanece en el medio y no se representa ha dejado perplejos a muchos comentaristas. Podemos darle sentido considerando la explicación de Simón (1969) de cómo una hormiga se mueve en su medio siguiendo un conjunto simple de rutinas y sin desarrollar [168-169] un mapa interno complejo. Puede haber detectores que determinen qué senda es más llana y un procedimiento que elige seguir esa senda. La hormiga responde, por tanto, a la información sobre los contornos de su medio pero no representa esos contornos. Pylyshyn (1979/1981) trata el ejemplo de Simón como algo que muestra cómo el control de un sistema puede estar parcialmente localizado en el medio en el que está el sistema sin que el sistema represente la información a sí mismo. (Ver Winograd, 1981, para un punto de vista relacionado.) Así pues, la información no tiene que representarse simbólicamente para ser útil a un sistema cognitivo. (Ver Dreyfus y Dreyfus, 1987, para objeciones adicionales al enfoque computacional basado en análisis de pericia humana.)

La objeción de Dreyfus, incluso si es válida, no arruina totalmente el Funcionalismo Computacional. Es posible admitir que alguna información está procesada en un sistema formal computacional y otra se halla en el medio o en prácticas aprendidas. El Funcionalismo Computacional podría con todo dar cuenta de ese procesamiento que incluye símbolos formales. Sin embargo, en la medida en que renunciamos a las explicaciones computacionales formales, bien sea porque consideramos que la información está en el medio o porque empleamos modelos no computacionales como el conexionismo, destruimos las pretensiones del Funcionalismo Computacional de dar una caracterización completa y general de la mente.

7.3.3. ESTADOS CUALITATIVOS Y LA OBJECIÓN DE LOS *QUALIA*

Una de las objeciones al Funcionalismo más ampliamente discutida es la afirmación de que no puede dar cuenta del carácter afectivo o cualitativo de los estados mentales. Se afirma que, cuando un computador se programa para identificar imágenes visuales, no experimenta la imagen, y que, cuando se programa para jugar al ajedrez, jamás siente inquietud alguna de ganar o perder. Podríamos intentar remediar este problema incorporando estados afectivos en nuestro análisis causal del sistema. Podríamos postular un homúnculo que reconozca cuándo es apropiado un cierto estado afectivo y altere el procesamiento del sistema de manera apropiada. Los críticos objetan, sin embargo, que, aunque tal estrategia podría proporcionar simulaciones más realistas de la conducta humana, las simulaciones resultantes no tendrían las experiencias que tiene un humano —no sentirán realmente dolor ni sufrirán

inquietud—.

Nagel (1974/1980) ha presentado este problema planteando la [169-170] pregunta «¿Qué es ser como un murciélago?». Podemos aprender con completo detalle cómo operan los mecanismos en el sistema de sonar de un murciélago, pero, con todo, no podemos imaginar cómo se sentirían cosas mediante el sonar. Esto es lo que deja de lado la explicación funcionalista, afirma él[10] (ver también Nagel, 1986). Jackson (1982) ha ofrecido un *Gedankenexperiment* para mostrarnos lo que el Funcionalismo no logra captar. Nos pide que contemplemos a una sofisticada neurofisióloga, María, que está privada de todas las experiencias de ver objetos coloreados pero que, con todo, desarrolla una explicación comprensiva de la operación del cerebro, incluyendo cómo realiza la percepción de colores. Aunque María sabe todo lo que hay que saber sobre los procesos cerebrales incluidos en la percepción de colores, aún no entiende la experiencia de ver rojo. Por consiguiente, concluye Jackson, el Funcionalismo no logra dar cuenta del carácter cualitativo de la vida mental.

Cierto número de filósofos ha intentado defender al Funcionalismo de estos ataques. Van Gulick (1986) ha mantenido que las propiedades afectivas serán propiedades funcionales de nivel superior. Ha mantenido también que, al conocer todo lo que hay que conocer sobre las propiedades de nivel inferior, podemos no conocer todavía las propiedades de nivel superior, y puede que tengamos que investigarlas separadamente. Pero, con todo, pueden ser todavía propiedades funcionales las que caracterizan cómo un sistema será capaz de interactuar con diversos tipos de fenómenos. Así, Van Gulick mantiene que los ejemplos avanzados por Nagel y Jackson están de acuerdo con el Funcionalismo y no se oponen a él.

Otra defensa se concentra en el hecho de que los argumentos de Nagel y de Jackson suponen que sabemos algo cuando tenemos un cierto género de experiencia. Es posible que no haya nada que conocer, sino sólo algo que experimentar[11]. Un modo relacionado de [170-171] presentar esta respuesta, debido a Lewis (1983a) y P.M. Churchland (1985), es afirmar que la palabra «sabe» se usa ambiguamente en las objeciones de Nagel y. de Jackson. Por una parte, se refiere al conocimiento conceptual y, por otra, a tener una experiencia. No hay razón para pensar que el conocimiento conceptual proporcione experiencia. P. S. Churchland (1986) ha ofrecido el ejemplo del embarazo, en el que surge una ambigüedad similar. Una ginecóloga que no tiene hijos puede conocer todos los procesos fisiológicos involucrados en el embarazo sin haber experimentado el haber estado embarazada. La ginecóloga sin hijos no carece de algo que podría conocerse conceptualmente. Simplemente no ha tenido una cierta experiencia.

Esas respuestas aceptan todas ellas la plausibilidad de las historias de Nagel y de Jackson, pero cuestionan cómo deberían interpretarse. P. S. Churchland ha cuestionado también si es posible conocerlo todo sobre un género de experiencia sin conocer cómo sería la experiencia: «¿Cómo puedo valorar lo que María (la neurofisióloga de la historieta de Jackson) sabe y entiende si sabe todo lo que hay que saber sobre el cerebro? Todo es mucho, y esto significa, con toda probabilidad, que María tiene una comprensión del cerebro radicalmente diferente y más profunda que cualquier cosa que podamos concebir en nuestros más salvajes vuelos de la fantasía» (1986, p. 332). Esta comprensión más profunda puede significar que María ya sabe cómo es la experiencia de ver rojo, de modo que el ejemplo de Jackson no podría surgir.

Estrechamente relacionadas con las objeciones de Nagel y Jackson al Funcionalismo están un conjunto de objeciones que afirman que el Funcionalismo no puede dar cuenta de los caracteres

cualitativos particulares de la experiencia, comúnmente reificados y a los que se hace referencia como *qualia*. Estás objeciones reposan en un conjunto de *Gedankenexperiments* que contemplan que nuestra experiencia cualitativa puede alterarse o desaparecer totalmente sin que ocurra ningún cambio en los procesos causales captados por el análisis funcionalista. Block y Fodor (1972/1980) han presentado uno de tales *Gedankenexperiments* que postula a alguien cuyos estados funcionales son idénticos a los nuestros pero que ve invertidos los colores del espectro visual o que siente los dolores como placenteros. (Esto es conocido como la condición de los *qualia invertidos*.) Block [171-172] y Fodor mantienen que el sentimiento de dolor es crítico por lo que respecta al estado mental de dolor, puesto que «nada sería una instancia del tipo "estado de dolor" a menos que se sintiese como un dolor» (Block y Fodor, 1972/1980, p. 244). Si esas situaciones pudiesen ocurrir[12], el Funcionalismo parecería hacer frente a una dificultad sería, pues mostraría que algo que es crítico para ciertos estados mentales no sería captado en las relaciones causales que figuran en los análisis funcionalistas.

Block y Fodor han propuesto un segundo *Gedankenexperiment* que contempla la existencia de un organismo con los mismos estados funcionales que nosotros pero que no tiene carácter cualitativo alguno asociado con sus estados funcionales. (Esto se conoce como la condición de los *qualia ausentes*.) Block (1978/1980) ha presentado este *Gedankenexperiment* más gráficamente proponiendo la existencia de robots en los que todas las interacciones causales que se encuentran en nosotros están realizadas pero donde no ocurre estado cualitativo alguno. Un ejemplo postula una cabeza homuncular humana en la que muchas personas en miniatura llevan a cabo las tareas postuladas en un análisis funcional de uno de nosotros. Otro incluye el que cada ciudadano de China sea responsable de un cuadrado particular de la tabla de máquina que caracteriza a uno de nosotros y ejecute esa tarea siempre que se le solicite de modo que toda la nación china se convierta en una simulación de uno de nosotros[13]. Block ha [172-173] sugerido, usando la frase de Nagel, que «existe una duda *prima facie* respecto de si hay algo consistente en ser como el sistema provisto de cabeza compuesta de homúnculos» (Block 1978/1980, p. 278). Así, él afirma que es posible satisfacer la teoría funcionalista sin tener estado cualitativo alguno. El Funcionalismo no puede, por tanto, dar cuenta de los estados cualitativos de los sistemas cognitivos.

Los funcionalistas que han respondido a los argumentos de los *qualia* invertidos y de los *qualia* ausentes han perseguido generalmente una de las dos estrategias siguientes. O bien han intentado evitar las objeciones tratando a los *qualia* mismos funcionalmente, o han mantenido que los *qualia* se deben en parte a la substancia física que realiza las funciones mentales y, por tanto, no son algo que el Funcionalismo está obligado a explicar. Discuto cada respuesta brevemente.

La primera estrategia ha sido seguida por Shoemaker (1975/1980) en un intento de responder al argumento de los *qualia* ausentes. Ha argumentado que los *qualia* pueden caracterizarse, al menos en parte, por su capacidad para causar creencias sobre ellos mismos, y que sin esta propiedad funcional no los conoceríamos, ni siquiera por introspección, y que los problemas sobre *qualia* no surgirían. El hecho de que conocemos esos estados cualitativos muestra, por tanto, que tienen reglas y propiedades funcionales y elimina la posibilidad de *qualia* totalmente ausentes. (Ver Block, 1980c, y Shoemaker, 1981, para una discusión adicional.) Con respecto al argumento de los *qualia* invertidos, Shoemaker adopta una posición más débil. Admite que, además de las propiedades funcionales, los *qualia* pueden tener otras propiedades. Si se intercambian, entonces resulta una situación de espectro invertido. Shoemaker ha afirmado, sin embargo, que admitir los espectros invertidos de esta manera no arruina en

modo alguno el Funcionalismo puesto que son las propiedades funcionales de los *qualia* las que de hecho usamos para distinguir objetos del mundo sobre bases cualitativas.

Los Churchland, sin embargo, adoptan una posición más fuerte argumentando que los criterios funcionales de los *qualia* son los que los definen. Si hay otros rasgos de los *qualia* que pudieran invertirse, [173-174] no serían importantes para lo que los *qualia* son. Si un rasgo que era parte de ver azul se convierte ahora en una parte de ver rojo, lo trataríamos ahora como una parte del *quale* rojo. Así pues, los Churchland mantienen que sólo los criterios funcionales son importantes y que jamás surgirán situaciones reales de *qualia* invertidos (Churchland y Churchland, 1981).

Algo que ha parecido hacer a los *qualia* difíciles de explicar para el Funcionalista es que parecen ser datos simples de la experiencia, algo que carece del género de complejidad que serviría para integrarlos en un análisis funcionalista. Dennett (1978d), sin embargo, ha argumentado que el carácter monádico de los *qualia* de dolor es ilusorio. La razón por la que la idea de que un computador sienta dolor ha parecido tan problemática consiste en que hemos malinterpretado dolor como una propiedad cualitativa simple. Dennett apela a los diferentes efectos que distintos anestésicos y analgésicos tienen sobre el dolor para mostrar que realmente hay diferentes aspectos del dolor. En una vena similar, Lycan (1987) ha propuesto un *Gedankenexperimenten* el que, en primer lugar, uno administra sucesivamente varios medicamentos para eliminar los diferentes aspectos de la experiencia de dolor hasta que no queda ninguno, y a continuación le da la vuelta al proceso para producir la experiencia global de dolor. Tal descomposición y recomposición del dolor proporcionaría crédito a la afirmación de que, después de todo, los estados cualitativos son realmente estados funcionales complejos, no simplemente estados monádicos. Una evidencia adicional a favor de la complejidad funcional de los *qualia* viene proporcionada por la capacidad de la gente de aprender a diferenciar *qualia* más finamente (p. ej., mediante entrenamiento estético)[14].

La segunda estrategia para responder a los argumentos de los *qualia* ausentes y los *qualia* invertidos es apelar a las estructuras físicas en las que se realizan los estados funcionales para dar cuenta de su carácter cualitativo. Esto elimina a los *qualia* de la lista de cosas de las que debe dar cuenta el Funcionalismo. Gunderson (1971) rotuló coloristamente los aspectos cualitativos de nuestros estados mentales como «propiedades resistentes de programas», sugiriendo que se deben a propiedades básicas del mecanismo en el que estaban realizadas [174-175] las propiedades programables; contrastaba, por tanto, las propiedades cualitativas con las propiedades receptivas de programas. Uno de los argumentos recientes más interesantes a favor de este enfoque se encuentra en Lewis (1980). Éste presenta dos casos hipotéticos que, aparentemente, nos presionan a hacer juicios inconsecuentes. El primero incluye a un marciano que está hecho de un mecanismo físico de un género diferente del nuestro (un sistema hidráulico) pero que tiene estados que son funcionalmente equivalentes a nuestros estados de dolor. El segundo incluye un loco que está en el mismo estado físico que aquél en el que estamos nosotros cuando sufrimos dolor pero que no muestra ninguno de los síntomas conductistas/funcionales de dolor (de hecho, siempre que esos estados ocurren, el loco se vuelca completamente en su trabajo y no hace nada para intentar referirse a ellos). Lewis encuentra que es intuitivo juzgar que ambos individuos tienen dolor, pero afirma que nuestros fundamentos para hacerlo en los dos casos son inconsecuentes. Al juzgar que el marciano ha sufrido dolor, empleamos criterios conductistas/funcionales (los estados del marciano desempeñan el mismo papel funcional que nuestros estados de dolor) mientras que al juzgar que el loco

tiene dolor estamos apelando al estado físico que realiza el dolor (el loco está en los mismos estados físicos que estamos nosotros cuando sufrimos dolor). Así pues, parece que estamos comprometidos tanto con un criterio funcional como con un criterio físico para identificar estados de dolor.

Para reconciliar estos criterios que están aparentemente en contradicción, Lewis ha propuesto que usemos un criterio funcional para determinar qué dolor es relativo a una especie, pero que usemos un criterio físico dentro de la especie. Así, un género de estado es un estado de dolor para miembros de una especie si es el género de estado que en los miembros normales de la especie funciona similarmente al modo en que los estados de dolor funcionan en los humanos. Dentro de una especie el dolor puede identificarse con el género de estado físico que en la mayor parte de los miembros de la especie instancia las relaciones causales que caracterizan funcionalmente dolor, incluso si no realiza esa función para un miembro particular de la especie. Lewis admite también que, si hay subgrupos distintos dentro de la especie en los que un mecanismo diferente realiza la función de dolor, podemos también usar ese mecanismo para valorar dolor en el subgrupo. Lewis (1980) aplica el mismo principio a inversiones del espectro de color:

> Yo diría que hay un buen sentido en el que la pretendida victima de los espectros invertidos ve rojo cuando mira al césped: está en un estado que [175-176] ocupa el papel de ver rojo para la humanidad en general. Y hay igualmente un buen sentido en el que ve verde: está en un estado que ocupa el papel de ver verde para él y para un pequeño subgrupo de población del que él es un miembro intachable y que tiene algún derecho a ser considerado como un género natural. Se tiene derecho a decir cualquiera de las dos cosas, aunque no a la vez. ¿Se necesita decir más? (p. 220).

La estrategia de Lewis, sin embargo, puede que no tenga éxito total. Lycan (1987), ha introducido dos casos adicionales en los que el análisis de Lewis parece dar resultados inaceptables. Uno incluye considerar el material físico que cumple normalmente en nosotros el papel funcional de un *quale* de dolor y hacer que desempeñe en nosotros otro papel distinto. El enfoque de Lewis parecería estar comprometido a decir que sentimos dolor siempre que este material está en el mismo estado en el que está cuando representa su papel causal normal. El segundo incluye la instalación de un órgano artificial que funcionase del mismo modo que el órgano real del dolor. En este caso, Lewis parecería estar comprometido con la negación del *quale*, puesto que la persona carece del estado material que usualmente produce el dolor en nuestra especie. Los problemas que Lycan ha planteado se deben al hecho de que Lewis ha intentado reconciliar criterios inconsecuentes para dolor. Un modo de resolver esta dificultad es insistir simplemente en los aspectos físicos del estado cuando se da cuenta del carácter cualitativo distintivo de los estados mentales. Hacer totalmente de los *qualia* un asunto de la constitución física de una entidad parece, sin embargo, problemático. Si la persona no puede diferenciar los estados cualitativamente, parece erróneo mantener que la persona en cuestión ha experimentado estados diferentes. La alternativa es volver a la primera estrategia y, como Churchland, Dennett y Lycan, adoptar un criterio totalmente funcionalista.

Aunque ambas estrategias parecen mantener su promesa de resolver el problema de los *qualia*, a muchos les sigue pareciendo un problema molesto. Parece que hay algunos aspectos de la experiencia que están más allá de lo que puede captarse en los modelos mecánicos de operación de la mente del

Funcionalismo. Esto nos retrotrae a donde comencé esta objeción al Funcionalismo con la preocupación de Nagel de que los análisis mecánicos jamás pueden captar el sentido en el que hay algo que es ser como un cierto género de sistema cognitivo. Como resultado de esto, la cuestión de los *qualia* continúa siendo uno de los temas más discutidos de la literatura funcionalista[15]. [176-177]

7.3.4. OBJECIONES DE CHAUVINISMO Y LIBERALIDAD

En algunos aspectos el Funcionalismo parece definir un campo equidistante entre el conductismo filosófico y la Teoría de la Identidad. AI igual que el conductismo filosófico, apela a criterios conductistas para caracterizar los fenómenos mentales, pero, a diferencia de él, el Funcionalismo interpreta los estados mentales como estados internos y les otorga un papel causal en la producción de la conducta. Al apoyar los estados mentales como procesos internos, el Funcionalismo está de acuerdo con la Teoría de la Identidad, pero difiere en que no insiste en qué tipos de estados mentales se identifican con estados del cerebro. Una de las críticas más interesantes del Funcionalismo, debida a Block (1978/1980), es que este campo medio es insostenible, y que el Funcionalismo tiene que sucumbir o bien ante un problema al que se enfrenta el conductismo filosófico o a un problema al que se enfrenta la Teoría de la Identidad. O el Funcionalismo será como el conductismo filosófico al ser demasiado liberal atribuyendo estados mentales a sistemas a los que no deberían atribuirse, o será, como la Teoría de la Identidad, demasiado chauvinista al negar estados mentales a sistemas que los tienen. A qué problema sucumbirá el Funcionalismo depende, de acuerdo con Block, de la forma de Funcionalismo que se adopte.

Block mantiene que el Funcionalismo de la psicología popular será demasiado liberal. Lo mismo que el conductismo filosófico atribuye estados mentales a cualquier sistema que tenga disposiciones conductistas apropiadas, el Funcionalismo de la Psicología Popular atribuye estados mentales, *qualia* y todo lo demás, a cualquier sistema que podamos caracterizar en términos de psicología popular. Si, por ejemplo, la nación china llevase a cabo una simulación de las interacciones causales que ocurren en mí tendríamos que atribuirle los mismos estados mentales que ahora se me atribuyen. En particular, afirma Block, el Funcionalista tiene que mantener que experimenta el mismo tipo de *qualia*. Esto, afirma Block, sería demasiado liberal, puesto que parece absurdo pensar que esta entidad compuesta tuviese [177-178] estados mentales, especialmente estados mentales cualitativos[16].

La forma alternativa de Funcionalismo que Block toma en consideración es lo que él llama «Psicofuncionalismo». Corresponde, de una manera amplia, a las tres versiones del Funcionalismo distintas del Funcionalismo de la Psicología Popular que se introdujeron al principio de este capítulo, donde los *procesos* causales incluidos en el análisis del funcionalista son los postulados en diversas teorías psicológicas o neurofisiológicas. Block mantiene que el psicofuncionalismo evita la objeción de ser demasiado liberal, puesto que elimina la atribución de estados mentales a cualquier sistema que no use los mismos procesos que producen en nosotros estados mentales[17]. Sin embargo, Block continúa argumentando que el psicofuncionalismo, al igual que la Teoría de la Identidad como Tipo, es demasiado chauvinista puesto que no nos permite atribuir estados mentales a organismos a los que deberíamos atribuírselos. Por ejemplo, no podríamos atribuírselos a marcianos que podrían vivir en gran medida de la manera en que nosotros vivimos, si bien usando procesos causales internos diferentes. Pero, mantiene

Block, deberíamos ser capaces de atribuir estados psicológicos a tales organismos si su conducta es apropiada: «seguramente hay muchas maneras de encajar en la descripción de la diferencia marciano-terrícola que he bosquejado de acuerdo con la cual sería perfectamente claro que, incluso si los marcianos se comportan diferentemente de nosotros en sutiles experimentos psicológicos, sin embargo ellos piensan, desean, experimentan alegría, etc. Suponer otra cosa sería un crudo chauvinismo humano» (Block, 1878/1980, p. 292).

Parecería que podemos dar cabida a los marcianos si sus propiedades causales fueran similares a las nuestras, incluso si fueran exactamente las mismas. Block mantiene que cualquier relajación de la exigencia de ser como nosotros nos convertiría en demasiado liberales. [178-179] Para evitar el exceso de liberalismo necesitamos especificar límites respecto de qué género de sistema es suficientemente similar a nosotros para permitirnos atribuir estados mentales, pero al imponer tales límites corremos el riesgo de ser demasiado chauvinistas. Block mantiene entonces que el Funcionalismo no tiene escapatoria: o ser demasiado liberal o ser demasiado chauvinista.

Para subrayar la seriedad de este problema, Block se ha centrado en un caso especial. En cualquier análisis funcional deben especificarse los *inputs* y *outputs* causales del sistema. Block afirma que el Funcionalismo no puede hacer esto sin ser demasiado liberal o demasiado chauvinista. Podemos intentar caracterizar *inputs* y *outputs* funcionalmente en términos de cualquier cosa que sucede para proporcionar el input y ser el output de un sistema, pero esto es, con mucho, demasiado liberal. Para mostrar esto, Block imagina un caso en el que unos manipuladores financieros podrían dirigir la economía boliviana de modo que esto instancie las relaciones funcionales que se encuentran en nosotros. Si caracterizamos sus *inputs* y *outputs* como cualquier cosa que induce procesos causales, estamos comprometidos con la afirmación de que la economía boliviana posee los mismos estados mentales que nosotros tenemos. Pero, comenta él, «si hay algún punto seguro al discutir el problema mente-cuerpo, uno de ellos es que la economía de Bolivia no puede tener estados mentales, sin importar lo distorsionada que esté por poderosos intrigantes» (p. 294). Pero, si exigimos que el sistema responda al mismo género de *inputs* físicos a los que nosotros respondemos y produzca los mismos *outputs*, somos una vez más chauvinistas, negando estados mentales a sistemas cognitivos posibles que tratan con diferentes *inputs* y *outputs*. Sin una explicación de *inputs* y *outputs* que evite tanto el chauvinismo como la liberalidad, Block afirma que es imposible caracterizar sistemas mentales en términos funcionales, esto es, en términos de relaciones causales entre tales estados e *inputs* y *outputs*.

Block puede haber identificado una limitación real de las versiones del Funcionalismo discutidas hasta ahora. Proporcionar una base para decidir qué clase de sistemas causalmente interactivos poseen estados mentales puede exigirnos considerar para qué propósitos sirven los procesos (Richardson, 1979). El invocar propósitos en un análisis funcional nos fuerza a una perspectiva teleológica. Aunque mucha gente piensa que una perspectiva teleológica es incompatible con la ciencia natural, sin embargo cierto número de filósofos de la biología recientes ha intentado mostrar cómo puede incorporarse una perspectiva teleológica dentro de la ciencia natural. En la sección siguiente bosquejo tal análisis. [179-180]

7.4 UNA VERSIÓN TELEOLÓGICA DEL FUNCIONALISMO

Los instrumentos básicos para un análisis teleológico de enunciados de función se introdujeron en el

capítulo 4, donde bosquejé cómo podría desarrollarse el análisis de Dennett de la intencionalidad dentro de una armazón evolucionista. Lo crítico en este asunto era tratar los estados mentales como rasgos adaptativos de organismos e interpretarlos en términos de los rasgos del medio con los que el organismo ha de habérselas para sobrevivir. Esta apelación a una armazón evolucionista nos permite también desarrollar un análisis teleológico general de función. La estrategia básica fue desarrollada por filósofos de la biología tales como Wright (1976) y Wimsatt (1972). Ambos apelaron al hecho de que, si una especie ha sido seleccionada puesto que poseía un rasgo particular, entonces ese rasgo servía para una necesidad de los miembros de la especie. Además, la presencia del rasgo en los miembros corrientes de la especie puede explicarse apelando a cómo ello capacita a la especie para cumplir esas presiones de la selección. Wright y Wimsatt mantenían, por tanto, que podemos atribuir al rasgo la función de servir a las necesidades de la especie.

Wright (1976, p. 81) ha ofrecido la siguiente especificación formal de cuándo es apropiado atribuir una función particular a alguna entidad:

> La función de *X* es *Z* si y sólo si:
>
> i) *Z* es una consecuencia (resultado) de que *X* exista, y
>
> ii) *X* existe porque hace (resulta en) *Z*.

La caracterización de la función por parte de Wright parece apoyar la causación hacia atrás puesto que es lo que *X* hace lo que se considera la causa de la ocurrencia de *X*. Pero esto no es el caso. Cuando se contempla desde una perspectiva evolucionista, la *X* de la cláusula ii) se refiere a una instancia del género *X* que es descendiente de otra instancia, y es a esa instancia anterior a la que se hace referencia en la cláusula i). Es esta instancia anterior del tipo la que tenía la consecuencia benéfica, y el hecho de tener esa consecuencia benéfica es lo que ha ocasionado la actual instancia. Así pues, no hay implicado nada más que la causación ordinaria. El análisis de Wimsatt de las funciones es un poco más elaborado, pero saca a la luz la relevancia de factores tales como la naturaleza del sistema y del medio, así como la propia perspectiva teórica de uno mismo a la hora de atribuir funciones. De este modo, él propone analizar las atribuciones de función en términos del esquema: «De acuerdo con la teoría *T* una función [180-181] de conducta *C* del elemento *e* en el sistema *S* en el medio *M* relativa al propósito *P* es hacer *Q*» (Wimsatt, 1972, p. 32). Este análisis se convierte en teleológico mediante el propósito y los valores de la teoría. El propósito viene dado por los factores de selección que gobiernan un sistema, y la teoría especifica los criterios que el sistema tiene que satisfacer para ser seleccionado y cómo la conducta del elemento satisface esos criterios.

Las explicaciones de Wright y Wimsatt tienen la virtud de invocar la función para la que sirve algo en la explicación de la ocurrencia actual de esa entidad usando sólo la causación eficiente. Esto permite la introducción de una perspectiva teleológica sin violar un punto de vista mecánico de la naturaleza. Hay, sin embargo, dos objeciones serias a este enfoque. En primer lugar, las posiciones de Wright y Wimsatt entrañan que algo que emerge sin una historia evolucionista pero que cumple las necesidades de un sistema no puede ser funcional. Esto es contraintuitivo. Si aceptamos, de acuerdo con la sabiduría

popular, que las jirafas adquirieron sus largos cuellos a causa de las ventajas que obtenían para conseguir comida, entonces, aunque podríamos decir que la función del cuello de la jirafa es ayudar para adquirir comida, no podríamos decir la misma cosa de una jirafa producida artificialmente por medio de ingeniería genética, puesto que carece de esta historia evolucionista. Pero su largo cuello la capacita también para cumplir las exigencias de reproducción y, de este modo, parece que serviría para esta función. (Este ejemplo se debe a Burian, comunicación personal, diciembre de 1983. Ver Margolis, 1976, para un ejemplo similar, y Short, 1983, para argumentos en contra de estos ejemplos.) En segundo lugar, los análisis de Wright y Wimsatt entrañan también que los órganos vestigiales que ayudaron a los primeros miembros de una especie a hacer frente a las exigencias del medio sirven aún para su función incluso si las exigencias del medio no están ya operando. Sus análisis parecen comprometernos a contar como funcional el gene de la anemia de células en forma de hoz debido a la protección que proporcionaba contra la malaria, aunque la malaria ya no presenta una fuerza de selección para los portadores de célula en forma de hoz, y ser un portador de célula en forma de hoz es un inconveniente, no una ventaja.

Hay una manera simple de remediar esos problemas. Más bien que exigir que las funciones sean adaptaciones (esto es: producto de la selección), necesitamos exigir que sean adaptativas (esto es: que incrementen la probabilidad de que el organismo se *reproducirá*). (Esta distinción se debe a Brandon, 1981.) Es decir, al averiguar qué es la función de algo, deberíamos mirar como beneficiará el rasgo al organismo en cuestión en su lucha por la supervivencia en lugar de [181-182] mirar cómo ha ayudado a sus antecesores. Hay, sin embargo, un coste significativo para este remedio. En la medida en que no estamos apelando al origen de un rasgo al adscribirle una función, no estamos explicando su ocurrencia y no deberíamos hablar de «explicaciones funcionales», sino sólo de «análisis funcionales» (ver Bechtel, 1986)[18].

Al invocar esta suerte de análisis funcional, podemos vencer las objeciones que Block ha planteado contra las versiones no ideológicas del Funcionalismo. Lo que la perspectiva teleológica nos exige hacer no es simplemente considerar las interacciones causales al identificar funciones, sino considerar cómo esos procesos causales están contribuyendo a las *necesidades* del organismo, tal como están especificadas por exigencias del medio[19]. Si un proceso no está contribuyendo al intento del organismo de hacer frente a las fuerzas de selección que operan sobre él, no se interpretará como una función. Considérese el ejemplo de Block de la nación china. Cuando los chinos simulan mis estados mentales, no están haciéndolo para cumplir con el mismo género de fuerzas de selección que las que actúan sobre mí. Por tanto, no necesitamos atribuir a los chinos mis estados mentales. Los chinos no constituyen un sistema que interactúa con un medio del género correcto. No están en el negocio de procesar estímulos sensoriales sobre los objetos ordinarios a los que se enfrenta una persona en la vida y planear acciones como respuesta. Constituyen un sistema social y, si llevasen a cabo el género de simulación que Block tiene en la cabeza, la fuerza de selección a la que estarían respondiendo sería la necesidad de ingresos y de prestigio como nación. Incluso aquí tenemos difícil la identificación del sistema en cuestión de una manera apropiada para un sistema evolucionista, puesto que [182-183] las naciones muy extensas pueden no tener el género de cohesión y continuidad que tienen los organismos. El sistema económico de Bolivia, para considerar otro de los ejemplos de Block, parece estar bastante bien delimitado y tener una cohesión duradera. Sin embargo, puede interpretarse como algo que evoluciona frente a las presiones de la

selección. Pero aquí los géneros de presiones de la selección son tan radicalmente diferentes de aquellos a los que hace frente una persona, que la atribución de estados mentales a procesos dentro de la economía boliviana es, obviamente, un error.

Block podría responder muy bien a estas sugerencias afirmando que aún deben hacer frente a la objeción de ser demasiado liberales o demasiado chauvinistas. Este enfoque teleológico nos exige especificar el tipo de fuerzas de selección a las que un sistema debe responder para que los procesos que están dentro de él cuenten como mentales, y Block podría argumentar que esto es imposible. Contrariamente a lo que Block piensa, hay, sin embargo, alguna esperanza de éxito, aunque no podamos producir ahora los análisis apropiados. Para hacerlo necesitaríamos clarificar qué fuerzas de selección del medio son importantes para determinar el futuro de los sistemas cognitivos y desarrollar una explicación de qué sistemas adaptativos se caracterizan apropiadamente como mentales. Los etólogos y los teóricos evolucionistas están en el buen camino para caracterizar los géneros de actividades que los organismos necesitan llevar a cabo para sobrevivir en una gran variedad de medios, y los principios generales de los procesos evolucionistas en este nivel pueden estar en camino. Mayr (1974), por ejemplo, ha distinguido entre sistemas cerrados, que descansan sobre instintos, y sistemas abiertos, que pueden aprender. Esto proporciona una dicotomía útil a la hora de diferenciar estrategias de supervivencia. Los atributos psicológicos parecerían sólo ser aplicables a aquellos organismos que adoptan una estrategia abierta y aprenden qué conductas realizar. Tales sistemas tienen que ser sensibles a la información sobre sus medios y ser capaces de procesar esta información para determinar las respuestas apropiadas. Esto sugiere que podríamos ser capaces de desarrollar una explicación general de los procesos mentales en términos de sus papeles en sistemas abiertos (p. ej. , como procesos que figuran en el procesamiento de información a partir de un medio que, a continuación, determina estrategias de acción).

Block podría plantear aún la queja chauvinista de que los procesos evolucionistas ha sido estudiados solamente en nuestra biosfera y que no sabemos cómo generalizarlos a un tipo de biosfera totalmente distinto. Tal queja no sería, sin embargo, peculiar de la psicología. Estamos igualmente inseguros respecto a cómo transferir los conceptos [183-184] biológicos a otra biosfera, más allá de su dominio interno. Si los principios fundamentales de nuestras ciencias biológicas y psicológicas pudiesen adaptarse al nuevo contexto de modo que nos diesen información útil probablemente los extenderíamos así y, a continuación, expandiríamos nuestra concepción de la vida y de la mente en el mismo sentido. Si eso no fuera posible, buscaríamos presumiblemente una armazón distinta en la que describir y explicar los fenómenos nuevos que hemos descubierto.

Podría parecer que un análisis ideológico como éste elimina a artefactos tales como los computadores como candidatos a estados mentales. No es obvio que ellos evolucionen del modo en que lo hacen los organismos vivientes. Mantengo, sin embargo, que este análisis da la respuesta correcta para juzgar la cuestión de si los computadores tienen mentes. Los artefactos están constreñidos por fuerzas de selección. A menudo esas fuerzas operan sólo en la mente del diseñador, que invoca criterios para elegir qué sistema construir. Pero los computadores (como sistemas compuestos de *hardware* y *software*) pueden construirse de manera que sean capaces de adaptarse ellos mismos a lo largo del tiempo a las demandas de sus medios respectivos. Los programas que se modifican son un paso en esta dirección. El hecho de que los computadores contemporáneos no se ajusten estrechamente a un medio al que se están

adaptando hace problemático el atribuir estados mentales a tales sistemas. Sin embargo, no hay en principio obstáculos para crear sistemas de computadores que interactúen y se adapten mucho más íntimamente a las exigencias de su medio. Si atribuimos estados mentales a tales sistemas, será menos probable que seamos acusados de ser demasiado liberales.

En la discusión anterior de este capítulo presenté el Funcionalismo Homuncular sin tratarlo ideológicamente. De las distintas versiones del Funcionalismo ésta es, sin embargo, la que se interpreta más naturalmente de modo teleológico, y así ha sido caracterizado por sus proponentes principales, Dennett y Lycan. Como hemos visto, el Funcionalismo Homuncular empieza con una explicación de lo que lleva a cabo todo el sistema y a continuación intenta explicar esa realización descomponiendo ese sistema en subsistemas (homúnculos). La perspectiva ideológica entra en escena con la manera en la que especificamos las tareas que el sistema está realizando. Si hacemos eso usando expresiones idiomáticas intencionales y si adoptamos una perspectiva evolucionista sobre la intencionalidad, ya hemos introducido una perspectiva teleológica. Estamos tratando los estados mentales como estados adaptativos de organismos. Tal perspectiva es crítica, sin embargo, para el desarrollo de la explicación [184-185] homuncular. Hay muchos procesos causales que suceden en los organismos, y muy bien podríamos acabar explicando rasgos del organismo que no son realmente de interés. Sin especificar lo que el sistema está logrando mediante su procesamiento interno, careceríamos de guía respecto de qué rasgos del sistema deberíamos intentar explicar (Burge, 1982; Dennett, 1981a). Así pues, el Funcionalismo Teleológico es un complemento natural del Funcionalismo Homuncular.

El Funcionalismo Teleológico conlleva también la concepción filosófica de un análisis funcionalista mucho más próximo a la tradición del Funcionalismo en psicología. Como se ha observado al principio de este capítulo, la tradición psicológica del Funcionalismo, a diferencia de la tradición filosófica, adoptó una perspectiva evolucionista y miró a los procesos psicológicos en términos de su significación en el medio. Hay grandes diferencias entre los enfoques, por ejemplo, de James y de Skinner, pero comparten este punto de atención común: cómo las actividades de los organismos los convierten en adaptados a las exigencias de un medio. He observado también al principio que los funcionalistas filosóficos consideran que ellos mismos están dando análisis de los procesos mentales en tanto que caracterizados por el cognitivismo contemporáneo, una perspectiva que a muchos les parece que choca radicalmente con el Funcionalismo psicológico tal como se ejemplifica por el conductismo. La introducción de un elemento teleológico en el Funcionalismo filosófico sugiere, sin embargo, que la caracterización de los estados mentales en el cognitivismo puede reconciliarse con el aspecto funcionalista de movimientos como el conductismo. Los intentos de caracterizar como internos los procesos mentales no son inconsecuentes con intentar entender esos procesos en términos de cómo permiten que los organismos se comporten en sus medios respectivos (Ver Bechtel, en prensa, y Schnaiter, 1987). Por consiguiente, además de mostrarnos cómo responder a la objeción de chauvinismo al Funcionalismo por parte de Block, el Funcionalismo Teleológico abre la perspectiva de acercamiento entre la concentración del cognitivismo en el procesamiento interno y la concentración del conductismo en el medio.

7.5 RESUMEN

El Funcionalismo constituye en la actualidad el análisis dominante de los eventos mentales en filosofía de la mente. En este capítulo he pasado revista a diversas versiones preeminentes del Funcionalismo. He discutido también algunas de las principales objeciones que se han planteado en contra del Funcionalismo y las principales respuestas [185-186] funcionalistas. De entre las objeciones, la objeción de Block de que el Funcionalismo no puede evitar el dilema de ser o demasiado liberal o demasiado chauvinista al atribuir estados mentales era la que parecía mostrar claramente una limitación del Funcionalismo. Como respuesta a esa objeción introduje una versión teleológica del Funcionalismo que ha sido desarrollada dentro de la filosofía de la biología. He mostrado cómo el Funcionalismo Teleológico puede vencer la objeción de Block y, al hacerlo, poner al Funcionalismo filosófico más de acuerdo con la tradición del Funcionalismo en psicología.

[186-187]

POSDATA

En este volumen he intentado proporcionar una amplia introducción a los problemas de la filosofía de la mente y a las posiciones que los filósofos han tomado respecto de ellos. Como debe de estar claro, existen considerables desacuerdos sobre esos temas. Con todo, estos problemas son de importancia central para la ciencia cognitiva. Implícita o explícitamente, los científicos cognitivos deben tomar una posición sobre si se puede dar cuenta de la intencionalidad en términos naturalistas, o sobre cómo se relaciona la mente con el cerebro, o sobre cómo han de identificarse los eventos mentales. Este volumen ha intentado proporcionar una introducción suficiente a esos problemas y a los puntos de vista que se han avanzado de modo que otros científicos cognitivos puedan entrar activamente en la discusión. Se necesita, sin embargo, cierta cautela. Una vez que uno se pone a discutir esos problemas, tiene que admitir la responsabilidad por los puntos de vista que adopta. Como se ha visto a lo largo de este libro, los filósofos discrepan. Además, no son infalibles, de modo que ¡no se tome a los filósofos como autoridades últimas!

Hay otros temas relevantes para la filosofía de la mente que o no se han discutido o se han mencionado sólo brevemente en este texto. Dos de ellos, de particular relevancia, tienen que ver con el innatismo y con las imágenes mentales. Respecto del innatismo ha habido una discusión filosófica de amplio rango respecto de lo que significa para una capacidad cognitiva el ser innata y sobre qué capacidades son innatas de hecho (ver, p. ej. , Stich, 1979, y los artículos en Piatelli-Palmarini, 1980; y Block, 1980b). Con respecto a las imágenes mentales, hay discusiones sobre qué son las imágenes mentales y cómo podrían almacenarse en la cabeza (ver, p. ej., Anderson, 1978; Kosslyn, 1980; Pylyshyn, 1981; Smith y Kosslyn, 1981; y los artículos recogidos en Block, 1980b). Dos antologías que serán particularmente útiles para aquellos que busquen una perspectiva amplia sobre la filosofía de la mente actual son Block (1980a, 1980b) y Haugeland (1981a).

[187-188]

BIBLIOGRAFÍA

Muchos de los artículos que se relacionan en esta sección han aparecido en diversas antologías. A menudo he indicado la antología en la que se han recogido, así como el lugar original de publicación. Cuando se indican dos fechas para un artículo que está en esas circunstancias, la primera de ellas se refiere a la fecha original de la publicación y la segunda a la fecha de la antología. Las referencias de página que se dan en el texto aluden a las versiones de los artículos recogidos en antologías.

ABRAHAMSEN, A. A. (1987): «Bridging boundaries versus breaking boundaries: Psycholinguistics in perspectives *Synthese*, 72, 355-388.

AMUNDSON, R. (1987): *Two autonomous domains*, manuscrito no publicado.

ANDERSON, A. R. (1964): *Minds and machines*, Prentice-Hall, Englewood Cliffs, NJ. [Versión castellana de F. Martín, *Controversias sobre mentes y máquina*s, Tusquets, Barcelona, 1984.]

ANDERSON, J. R. (1978): «Arguments concerning representations for mental imagery», *Psychological Review*, 85, pp. 249-277.

ANDERSON, J. R.; GREENO , J. G.; KLINE , P. J., y NEVES , D. M. (1981): «Acquisition of problem solving skill», en J. R. Anderson (ed.), *Cognitive skills and their acquisition*, Lawrence Erlbaum Associates, Hillsdale NJ, pp. 191-230.

ANDERSON, R. E. (1986): «Cognitive explanations and cognitive ethology», en W. Bechtel (ed.), *Integrating scientific disciplines*, Reidel, Dordrecht, pp. 323-336.

ANSCOMBE, G. E. M. (1965): «The intentionality of sensation: A grammatical feature», en R. J. Butler (ed.), *Analytic philosophy*, Basil Blackwell, Oxford, 2.ª serie, pp. 158-180.

AQUILA, R. E. (1977): *Intentionality: A study of mental acts*, The Pennsylvania State University Press, University Park, PA.

ARMSTRONG, D. M. (1968): *A materialist theory of mind*, Routledge y Kegan Paul, London.

— (1980): *The nature of mind and other essays*, Cornell University Press, Ithaca, NY.

— (1984): «Consciousness and causality », en D. M. Armstrong y N. Malcolm (eds.), *Consciousness and causality: A debate on the nature of mind*, Basil Blackwell, Oxford, pp. 103-191.

AUSTIN, J. L. (1956-1957/1970): «A plea for excuses», *Proceedings of the Aristotelian Society*, 57, pp. 1-30. Reimpreso en J. O. Urmson y G. J. Warnock (eds.), *Philosophical papers of J. L. Austin*, Oxford University Press, Oxford, 1970, 2.ª serie, pp. 175-204. [Versión castellana: J. L. Austin, 1989.]

— (1962a): *How to do things with words*, J. O. Urmson (ed.), Oxford University Press, New York. [Versión castellana de G. R. Canio y E. A. Rabossi, *Cómo hacer cosas con palabras*, 2.ª ed., Paidós, Buenos Aires, 1988.]

— (1962b): *Sense and Sensibilia* (Reconstruido a partir de notas manuscritas por G. J. Warnock), Oxford University Press, Oxford. [Versión castellana en Tecnos, Madrid, 1981.]

— (1988): *Ensayos filosóficos*, versión castellana de Alfonso García Suárez, Alianza, Madrid.

BACON, F. (1620): *Novum organon*, J. Billium, London. [Versión castellana de C. Litrán, 2.ª ed., Orbis, Barcelona, 1985.]

BAILEY G. (1986): *Cognitive psychology and representational theories of mind*, manuscrito no publicado.

BARNETTE, R. L. (1977): «Kripke's pains», *Southern Journal of Philosophy*, 15, pp. 3-14.

BARSALOU, L. (en preparación): *Cognitive psychology: An overview for cognitive science*, Lawrence Erlbaum Associates, Hillsdale, NJ.

BEALER, G. (1978): «An inconsistency in functionalism», *Synthese*, 38, pp. 333-372.

BECHTEL, W. (1978): «Indeterminacy and intentionality: Quine's purported elimination of propositions» *Journal of Philosophy*, 75, pp. 649-662.

— (1980) «Indeterminacy and underdetermination: Are Quine's two theses consistent?», *Philosophical Studies*, 38, pp. 309-320.

— (1985a) «Realism, instrumentalism, and the intentional stance» *Cognitive Science*, 9, pp. 473-497.

— (1985b): «Attributing responsibility to computer systems», *Metaphilosophy*, 16, pp. 296-306.

— (1986): «Teleological functional analyses and the hierarchical organization of nature», en N. Rescher (ed.), *Teleology and natural science*, University Press of America, Landham, MD, pp. 26-48.

— (en prensa a): «Perspectives on mental models», *Behaviorism*, 17.

— (en prensa b): *Philosophy of science: An overview for cognitive science*, Lawrence Erlbaum Associates, Hillsdale, NJ.

— (en prensa c): «Connectionism and philosophy of mind: An overview», *Southern Journal of Philosophy*, 25.

BECHTEL, W. y RICHARDSON, R. C. (1983): «Consciousness and complexity: Evolutionary perspectives on the mind-body problem» *Australasian Journal of Philosophy*, 61, pp. 378-393.

BENNETT, J. (1976): *Linguistic behavio*r, Cambridge University Press, Cambridge.

BERKELEY, G. (1710/1965): «A treatise concerning the principles of human knowledge», en CM. Turbaye (ed.), *Principles, dialogues, and correspondence*, Bobbs-Merrill, Indianapolis, pp. 3-101.

BERNSTEIN, R. J. (1968/1971): «The challenge of scientific materialism», *International Philosophical Quarterly*, 8, pp. 252-275. Reimpreso en Rosenthal, 1971, pp. 200-222.

BIRO, J. I. (1985a): «Hume and cognitive science», *History of Philosophy Quarterly*, 2, pp. 257-274.

— (1985b, noviembre): *Kant and neuro-science*. Artículo presentado al XI Congreso Interamericano de Filosofía, Guadalajara.

BLOCK, N. (1978/1980): «Troubles with functionalism», en C. W. Savage (ed.), *Perception and cognition. Issues in the foundations of psychology. Minnesota studies in the philosophy of science*, University of Minnesota Press, Minneapolis, vol. 9, pp. 261-325. Reimpreso en Block, 1980, pp. 268-305.

— (1980a): *Readings in philosophy of psychology*, (vol. 1), Harvard University Press Cambridge, M. A.

— (1980b): *Readings in philosophy of psychology*, (vol. 2), Harvard University Press, Cambridge, M. A.

— (1980c): «Are absent qualia impossible?» *The Philosophical Review*, 89, pp. 257-274.

BLOCK, N., y FODOR J. A. (1972/1980): «What psychological states are not», *Philosophical Review*, 81, pp. 159-181. Reimpreso en Block, 1980, pp. 237-250.

BODEN, M. (1977): *Artificial intelligence and natural man*, Basic Books, New York, [Version castellana de J. C. Armero Sanjos6, Tecnos, Madrid, 1983.]

— (1981): *Minds and mechanisms: Philosophical psychology and computational models*, Cornell

University Press, Ithaca, NY.

BORST, C. V. (ed.) (1970): *The mind/brain identity theory*, Macmillan, New York.

BOVERI, T. (1903): «Über die Konstitution der chromatischen Kemsubstanz», *Verhandlungen der deutschen zoologischen gesellschaft zu Würzberg*, 13, pp. 10-33.

BOYD, R., y RICHERSON, P. J. (1985): *Culture and the evolutionary process*, University of Chicago Press, Chicago.

BRANDON, R. (1981): «Biological teleology: Questions and explanations», *Studies in the history and philosophy of science*, 12, pp. 91-105.

BRENTANO, F. (1973): *Psychology from an empirical standpoint*, trad, de A. C. Pancurello, D. B. Terrell y L. L. McAlister, Humanities, New York. (Obra publicada originalmente en 1874.)

BREWER W. F. (1974): «There is no convincing evidence for operant or classical conditioning in adult humans», en W. B. Weimer y D. S. Palermo (eds.), *Cognition and the symbolic process*, Lawrence Erlbaum Associates, Hilsdale, NJ, pp. 1-42.

BRICKE, J. (1984): «Dennett's eliminative arguments», *Philosophical Studies*, 45, pp. 413-429.

BÜRGE, T. (1979): «Individualism and the mental», *Midwest Studies in Philosophy*, 4, pp. 73-121.

— (1982): «Other bodies», en A. Woodfield (ed.), *Thought and object*, Oxford University Press, Oxford, pp. 97-120.

BYNUM, T. W. (1985): «Artificial intelligence, biology, and intentional states», *Metaphilosophy*, 16, pp. 355-377.

CAMPBELL, D. T. (1966): «Pattem matching as an essential in distal knowing», en K. R. Hammond (ed.), *The psychology of Egon Brunswik* Holt, Rinehart y Winston, New York, pp. 81-106.

CARLETON, L. R. (1984): «Programs, language understanding, and Searle», *Synthese*, 59, pp. 219-230.

CARNAP, R. (1956): *Meaning and necessity*, University of Chicago Press. Chicago.

COOK, T. D. y CAMPBELL, D. T. (1979): *Quasi-experimentation: Design and analysis for field settings*. Chicago: Rand McNally.

CORNMAN J. W. (1962): «Intentionality and intensionality», *Philosophical Quarterly*, 12, pp. 44-52.

— (1962-1971): «The identity of mind and body», *The Journal of Philosophy*, 59, pp. 486-492. Reimpreso en Rosenthal, 1971, pp. 73-79.

— (1968): «On the elimination of "sensations" and sensations», *The Review of Metaphysics*, 22, pp. 15-35.

— (1977): «Mind-body identity: Cross-categorial or not?», *Philosophical Studies*, 32, pp. 165-174.

CUMMINS, R.: (1975): «Functional analysis», *The Journal of Philosophy*, 72. pp. 741-760.

— (1983): *The nature of psychological explanation*, MIT Press/Bradford Books, Cambridge, M. A.

CHISHOLM, R. M. (1957): *Perceiving: A philosophical study*, Cornell University Press, Ithaca, NY.

— (1958): «Sentences about believing», en H. Feigl, M. Scriven y G. Maxwell (eds.), *Minnesota studies in the philosophy of science*. University of Minnesota Press, Minneapolis, MN, pp. 510-520.

— (1967): «Intentionality», en E. Edwards (ed.), *The encyclopedia of philosophy*, McMillan, New York, vol. 4, pp. 201-204.

— (1984): «The primacy of the intentional», *Synthese*, 61, pp. 89-109.

CHOMSKY, N.: (1959): «Review of Skinner's verbal behavior», *Language*, 35, pp. 26-58. [Versión castellana en R. Bayes (ed.), *¿Chomsky o Skinner? La génesis del lenguaje*, Fontanella, Barcelona, 1977.]

— (1966): *Cartesian linguistics: A chapter in the history of rationalist thought*, MIT. Press Cambridge, MA, [Versión castellana de E. Wulff, Linguistica cartesiana, Gredos, Madrid, 1984.]

— (1969): «Quine's empirical assumptions», en D. Davidson y J. Hintikka (eds.), *Words and objections. Essays on the work of W. V. Quine*, Reidel, Dordrecht, pp. 53-68.

— (1986): *Knowledge of language*, Praeger, New York, [Version castellana de E. Bustos Guadano, *Conocimiento del lenguaje*, Alianza, Madrid, 1989.]

CHURCH, A. (1943): «A review of Quine», *Journal of Symbolic Logic*, 8, pp. 45-47.

CHURCHLAND, P. M. (1979): *Scientific realism and the plasticity of mind*, Cambridge University Press, Cambridge.

— (1981a): «Eliminative materialism and propositional attitudes», *The Journal of Philosophy*, 78 pp. 67-90.

— (1981b). «Is *Thinker* a natural kind?» *Dialogue*, 21, pp. 223-238.

— (1984): *Matter and consciousness: A contemporary introduction to the philosophy of mind*, MIT Press/Bradford Books, Cambridge.

— (1985): «Reduction, qualia, and the direct introspection of brain states», *The Journal of Philosophy*, 82 pp. 8-28.

— (1986): «Some reductive strategies in cognitive neurobiology», *Mind*, 95, pp. 279-309.

CHURCHLAND, P. M. y CHURCHLAND, P. S. (1981): «Functionalism, qualia, and intentionality», *Philosophical Topics*, 12, pp. 121-145.

CHURCHLAND, P. S. (1978): «Fodor on language learning», *Synthese*, 38, pp. 149-159.

— (1980a): «Language, thought, and information processing», *Nous*, 14, pp. 147-170.

— (1980b): «A perspective on mind-brain research», *The Journal of Philosophy*, 11, pp. 185-207.

— (1983): «Consciousness: The transmutation of a concept», *Pacific Philosophical Quarterly*, 64, pp. 80-95.

— (1986): *Neurophilosophy: Toward a unified science of the mind-brain*, MIT Pres/ Bradford Books, Cambridge.

CHURCHLAND, P. S., y CHURCHLAND, P. M. (1983): «Stalking the wild epistemic engine», *Nous*, 17, pp. 5-18.

DARDEN, L., y MAULL, N. (1977): «Interfield theories», *Philosophy of Science*, 43, pp. 44-64.

DAVIDSON, D. (1967): «Truth and meaning», *Synthese*, 17, pp. 304-323. Reimpreso en D. Davidson. 1984, pp. 17-36. [Version castellana en L.M. Valdés Villanueva, 1991.]

— (1970/1980): «Mental events», en L. Foster y J. W. Swanson (eds.), *Experience and theory*, University of Massachusetts Press, Amherst, pp. 79-101. Reimpreso en Block, 1980, pp. 107-119.

— (1973): «Radical interpretation», *Dialectica*, 27 pp. 313-328. Reimpreso en Davidson, 1984, pp. 125-139. [Versión castellana en L. M. Valdés Villanueva, 1991.]

— (1974a): «Belief and the basis of meaning», *Synthese*, 21, pp. 309-323. Reimpresión en D. Davidson, 1984, pp. 141-154. [Versión castellana en Davidson, 1989.]

— (1974b): «On the very idea of a conceptual scheme», *Proceedings and Addresses of the American*

Philosophical Association, 47 pp. 5-20 Reimpreso en D. Davidson, 1984, pp. 183-198.

— (1975): «Thought and talk», en S. Guttenplan (ed.), *Mind and language*, Clarendon Press, Oxford, pp. 7-23. Reimpreso en D. Davidson, 1984, pp. 155-170.

— (1984): *Inquiries into truth and interpretation*, Clarendon Press, Oxford.

— (1989): *De la verdad y de la interpretación*, trad. de G. Filippi, Gedisa, Barcelona.

DENNETT, D. C. (1971/1978): «Intentional Systems», *The Journal of Philosophy*, 68, pp. 87-106. Reimpreso en D. C. Dennett, 1978a.

— (1975/1978):«Why the law of effect will not go away», *Journal of the Theory of Social Behavior*, 5, pp. 169-187. Reimpreso en D. C. Dennett, 1978a, pp. 71-89.

— (1977): «Critical notice of J. Fodor», *The Language of Thought*. *Mind*, 86, pp. 265-280.

— (1978a): *Brainstorms*, MIT Press/Bradford Books, Cambridge.

— (1978b): «Skinner skinned», en D. C. Dennett (ed.), *Brainstorms*, MIT Press/ Bradford Books, Cambridge, pp. 53-70.

— (1978c): «Toward a cognitive theory of consciouness» en D. C. Dennett (ed.), *Brainstorms*, MIT Press/Bradford Books, Cambridge, pp. 149-173.

— (1978d): «Why you can't make a computer that feels pain», D. C. Dennett (ed.), *Brainstorms*, MIT Press/Bradford Books, Cambridge, pp. 190-229.

— (1979): «Current issues in the philosophy of mind», *American Philosophical Quarterly*, 15, pp. 249-261.

— (1981a): «Three kinds of intentional psychology», en R. Healey (ed.) *Reduction, time and reality*, Cambridge University Press, Cambridge, pp. 37-61.

— (1981b): «Making sense of ourselves», *Philosophical Topics*, 12, pp. 63-81.

— (1981c): «True believers: The intentional strategy and why it works», en A. F. Heath (ed.), *Scientific explanation*, Clarendon Press, Oxford, pp. 53-75.

— (1982), «Beyond belief», en A. Woodfield (ed.), *Thought and object*, pp. 1-95.

— (1983): «Intentional systems in cognitive ethology: The "Panglossian paradigm" defended», *The Behavioral and Brain Sciences*, 6, pp. 343-390.

— (1984a): «Cognitive wheels: The frame problem of AI», en C. Hookway (ed.), *Minds, machines and evolution*, Cambridge University Press, Cambridge, pp. 129-151.

— (1984b): «I could not have done otherwise—so what?», *The Journal of Philosophy*, 81, pp. 553-565.

— (1984c): *Elbow room: The varieties of free will worth wanting*, MIT Press/Bradford Books, Cambridge, MA.

— (1986): «The logical geography of computational approaches. A view from the East Pole», en M. Brand y R. M. Harnish (eds.), *The representation of knowledge and belief*, University of Arizona Press, Tucson, pp. 59-79.

— (1987): «Consciousnes», en R. L. Gregory (ed.), *Oxford companion to the mind*. Oxford University Press, Oxford.

DESCARTES, R. (1637/1970): «Discourse on method», en E. S. Haldane y G. R. T. Ross (eds.), *The philosophical works of Descartes*, Cambridge University Press, Cambridge, vol. 1, pp. 11-151. [Versión castellana de E. Bello Reguera, *Discurso del método*, Tecnos, Madrid, 1987.]

— (1641/1970): «Meditations on first philosophy», en E. S. Haldane y G. R. T. Ross (eds.), *The philosophical works of Descartes*, Cambridge University Press, Cambridge, vol. 1, pp. 181-200. [Versión castellana de Vidal Peña, Alfaguara, Madrid, 1977.]

— (1644/1970): «Principies of philosophy», en E. S, Haldane y G. R.T. Ross (eds.), *The philosophical works of Descartes*, Cambridge University Press, Cambridge, vol. 1, pp. 178-291. [Versión castellana de F. Alcaide y Vilar, *Los principios de la filosofía*. Reus, Madrid, 1925.]

DONNELLAN, K. (1972): «Proper names and identifying descriptions», en D. Davidson y G. Harman (eds.), *Semantics of natural language*, Reidel, Dordrecht, pp. 356-379.

— (1974): «Speaking of nothing». *Philosophical Review*, 83, pp. 3-31.

DRETSKE, F. I. (1980): «The intentionality of cognitive states», *Midwest Studies in Philosophy*, 5, pp. 281-294.

— (1981): *Knowledge and the flow of information*, MIT Press/Bradford Books, Cambridge, MA. [Versión castellana en M. Vicedo Guilla, *Conocimiento y flujo de información*, Salvat, Pamplona, 1989.]

— (1983); «Precis of *Knowledge and the flow of information*», *The Behavioral and Brain Sciences*, 6, pp. 55-90.

DREYFUS, H. L. (1979): *What computers can't do: The limits of artificial intelligence*, 2nd ed., Harper and Row, New York.

— (1982): «Introduction», *Husserl, intentionality, and cognitive science*, MIT Press/ Bradford Books, Cambridge.

— (1985): *Artificial intelligence: The problem of knowledge representation*, manuscrito no publicado.

DREYFUS, H. L. y DREYFUS, S. E. (1987): *Mind over machine. The power of human intuition and expertise in the era of the computer*, The Free Press, New York.

ENC, B. (1983): «In defense of the identity theory», *Journal of Philosophy*, 80, pp. 279-298.

FALK, A. E. (1981): «Purpose, feedback, and evolution», *Philosophy of Science*, 48, pp. 198-217.

FEIGL, H. (1958/1967): *The «mental» and the «physical»: The essay and a postscript*, University of Minnesota Press, Minneapolis.

— (1960/1970): «Mind-body, *not* a pseudo problem», en S. Hook (ed.), *Dimensions of mind*, New York University Press, New York. Reimpreso en C. V. Borst, 1970, pp. 33-41.

FELDMAN, F. (1974): «Kripke on the identity theory», *Journal of Philosophy*, 71, pp. 665-676.

— (1980): «Identity, necessity, and events», en N. Block (ed.), *Readings in philosophy of psychology*, Harvard University Press, Cambridge, MA, vol. 1, pp. 148-155.

FEYERABEND, P. K. (1963/1970): «Materialism and the mind-body problem», *The Review of Metaphysics*, 17, pp. 49-67. Reimpreso en C. V. Borst, 1970.

FIELD, H. H. (1978/1980): «Mental representation», *Erkenntnis*, 13, pp. 9-61. Reimpreso en C. V. Borst, 1980.

FODOR, J. A. (1968): *Psychological explanation*, Random House, New York. [Versión castellana de J. E. García Albea, *La explicación psicológica*, Cátedra, Madrid, 1980.]

— (1974): «Special sciences (Or: Disunity of science as a working hypothesis)», *Synthese*, 28, pp. 97-115.

— (1975): *The language of thought*, Crowell, New York. [Versión castellana de J. Fernandez

Zulaica, El lenguaje del pensamiento, Alianza, Madrid, 1985.]

— (1980): «Methodological solipsism considered as a research strategy in cognitive psychology», *The Behavioral and Brain Sciences*, 3, pp. 63-109. Reimpreso en J. Haugeland, 1981.

— (1981): «The present status of the innateness controversy», en J. A. Fodor (ed.), *Representations*, MIT Press/Bradford Books. Cambridge, pp. 257-316.

— (1983): *The modularity of mind*, MIT Press/Bradford Books, Cambridge MA [Versión castellana de J. M. Igoa, *La modularidad de la mente*, Morata, Madrid, 1986.]

— (1984): «Semantics, Wisconsin style», *Synthese*, 59, pp. 231-350.

— (1985): «Precis of *The modularity of mind*», *The Behavioral and Brain Sciences*, 8, pp. 1-42.

— (1987): *Psychosemantics: The problem of meaning in the philosophy of mind*, MIT Press, Cambridge, MA.

FODOR, J. A. y PYLYSHYN, Z. W. (1981): «How direct is visual perception? Some reflection on Gibson's "ecological approach"», *Cognition*, 9, pp. 136-196.

— (1987): *Connectionism and cognitive architecture: A critical analysis*, manuscrito no publicado.

FOLLESDAL, D. (1982): «Brentano and Husserl on intentional objects and perception», en H. L. Dreyfus (ed.). *Husserl, intentionality, and cognitive science*, MIT Press/Bradford Books, Cambridge, MA, pp. 31-41.H

FREGE, G. (1982): «Über Sinn and Bedeutung», *Zeitschrift fur Philosophie und philosophised Kritik*, 100, pp. 25-50. [Versión castellana en L. M. Valdés Villanueva, 1991.]

FURTH, H. (1966): *Thinking without language: Psychological implications of deafness*, The Free Press. New York. [Versión castellana de R. Martínez Arias, *Pensamiento sin lenguaje*, Marova, Madrid, 1981.]

GARDNER, R. A., y GARDNER, B. T. (1969): «Teaching sing language to a chimpanzee», *Science*, 165, pp. 664-672.

GASSENDI, P. (1641/1970): «Letter to Descartes. In "Objections and replies"», en E. S. Haldane y G. R. T. Ross (eds.), *The philosophical works of Descartes*, Cambridge University Press. Cambridge, vol. 2, pp. 179-240.

GAUKER, C. (1987): *Thought as inner speech*, manuscrito no publicado.

GEACH, P. T. (1957): *Mental acts*, Routledge and Kegan Paul, London.

GIBSON, J. J. (1979): *The ecological approach to perception*, Houghton Mifflin, Boston.

GLOTZBACH, P., y HEFT, H. (1982): «Ecological and phenomenological contributions to the phenomenology of perception», *Nous*, 16, pp. 108-121.

GOODMAN, N. (1955): *Fact, fiction, and forecast*, Harvard University Press, Cambridge, MA.

GOULD, S. J., y LEWONTIN, R. C. (1979): «The spandrels of San Marco and the panglossian paradigm: A critique of the adaptationist programme», *Proceedings of the Royal Society of London*, B205, pp. 581-598.

GREEN, G. (en preparación): *Linguistic Pragmatics for Cognitive Science*, Lawrence Erlbaum Associates, Hillsdale, NJ.

GRICE , H. P. (1975): «Logic and conversation», en P. Cole y J. L. Morgan (eds.), *Speech acts*, Academic Press, New York, pp. 45-58. [Versión castellana en L. M. Valdés Villanueva, 1991.]

GUNDERSON , K. (1970/1971): «Asymmetries and mind-body perplexities», en M. Radner y S.

Winokur (eds.), *Minnesota studies in the philosophy of science*, University of Minnesota Press, Minneapolis, vol. 4, pp. 273-309. Reimpreso en D.M. Rosenthal, 1971, pp. 112-127.

— (1971): *Mentality and machines*, Anchor Books, Garden City, NY.

HAMILTON, E., y CAIRNS, H. (1961): *The collected dialogues of Plato*, Bollingen, New York.

HAMLYN, D. W. (1977): «The Concept of information in Gibson's Theory of Perception», *Journal for the Theory of Social Behavior*, 7, pp. 5-16.

HARDIN, C. L. (1985): *Qualia and materialism: Closing the explanatory gap*, manuscrito no publicado.

— (1988): *Color for philosophers*. Hacket Publishing, Indianapolis.

HARMAN, G. (1973): *Thought*, Princeton University Press, Princeton.

— (1977): «How to use propositions», *American Philosophical Quarterly*, 14, pp. 171-176.

— (1978): «Is there mental representation?», en C. Wade Savage (ed.), *Minnesota studies in the philosophy of science. Perception and cognition — Issues in the foundation of psychology*. University of Minnesota Press, Minneapolis, vol. 9, pp. 57-63.

HARNAD, S. (1987): *Minds, machines and Searle*, manuscrito no publicado.

HAROUTUNIAN, S. (1983): *The equilibrium model of explanation: Strengtiis and limitations for an account of cognitive change*, Springer-Verlag, New York.

HATFIELD, G. (1986): *Representation and content in some (actual) theories of perception*, Reports of the Cognitive Neuropsychology Laboratory, Johns Hopkins University, n.° 21.

HAUGELAND, J. (1981a): *Mind design*, MIT Press/Bradford Books, Cambridge, MA.

— (1981b): «Semantic engines: An introduction to mind design», en J. Haugeland (ed.). *Mind design*, MIT Press/Bradford Books, Cambridge, MA, pp. 1-34.

— (1985): *Artificial intelligence: The very idea*, MIT Press/Bradford Books, Cambridge, MA.

HEIDEGGER, M. (1962): *Being and time*, Harper and Row, New York. Originalmente publicado en 1949. [Versión castellana de J. Gaos, *El ser y el tiempo*,7.ª ed., Fondo de Cultura Económica, México, 1989.]

HEIL, J. (1983): *Perception and cognition*, University of California Press, Berkeley.

HILL, C. S. (1984): «In defense of type materialism». *Synthese*, 59, pp. 295-320.

HORGAN, T. (1982): «Jackson on physical information and qualia», *The Philosophical Quarterly*, 32, pp. 147-156.

— (1984): «Functionalism, qualia, and the inverted spectrum», *Philosophy and phenomenological research*, 44, pp. 453-469. HUME, D. (1748/1962): Enquiry concerning the human understanding, Clarendon Press, Oxford. [Versión castellana de J. Salas Ortueta, *Investigación sobre el conocimiento humano*, 6.ª ed.. Alianza. Madrid, 1988.]

— (1759/1888): *A treatise of human nature*, Clarendon, Oxford. [Versión castellana de F. Duque, Tratado de la naturaleza humana, Tecnos, Madrid, 1988.]

HUSSERL, E. (1913/1970): *Logische Untersuchungen,* Niemeyer, Halle. [Versión castellana de M. García Morente, *Investigaciones lógicas*, 2 vols., 2.ª ed., Alianza, Madrid, 1985.]

— (1929/1960): *Cartesian meditations*. Nijhoff, The Hague. [Versión castellana de M. A. Presas, Meditaciones cartesianas. Tecnos, Madrid, 1986.]

— (1950/1972): *Ideen zu einer reinen Phänomenologie and phänomenologischen Philosophie. I,*

Husserliana, Nijhoff. The Hague. [Versión castellana de J. Gaos, *Ideas relativas a una fenomenología pura y una filosofía fenomenológica*, Fondo de Cultura Económica, México, 1985.]

JACKSON, F. (1982): «Epiphenomenal qualia», *Philosophical Quarterly*, 32, pp. 127-136.

JACOBY, H. (1985): «Eliminativism, meaning, and qualitative states», Philosophical Studies, 47, pp. 257-270.

KAHNEMAN, D.; SLOVIC, P., y TVERSKY, A. (1982): *Judgment under uncertainty: Heuristics and biases*, Cambridge University Press, Cambridge.

KANT, I. (1787/1961): *Critique of pure reason*, trad, de N. K. Smith, MacMillan, London. [Versión castellana de P. Ribas, *Critica de la razón pura*, 7. ª ed., Alfaguara, Madrid, 1989.]

KAPLAN, D. (1967): *Transworld identification*, articulo presentado en la Western Division, American Philosophical Association, abril, Chicago, IL.

— (1969): «Quantifying in», en D. Davidson y J. Hintikka (eds.), *Words and objections: Essays on the work of W. V. Quine*, Reidel, Dordrecht, pp. 206-242.

— (1978): «Dthat», en Peter Cole (ed.), *Syntax and semantics*, Academic Press, New York, vol. 9, pp. 221-243.

KENNY, A. (1970): *Descartes' philosophical letters*, Clarendon Press, Oxford.

KIM, J. (1978): «Supervenience and nomological incommensurables», *American Philosophical Quarterly*, 15, pp. 149-156. 4

— (1979): «Causality, identity, and supervenience in the mind-body problem», *Midwest Studies in Philosophy*, 4, pp. 31-49.

— (1982a): «Psychophysical supervenience as a mind-body theory», *Cognition and Brain Theory*, 5, pp. 129-147.

— (1982b): «Psychophysical supervenience», *Philosophical Studies*, 41, pp. 51-70.

— (1985): «Psychophysical laws», en E. Lepore y B. McLaughlin (eds.), *Actions and events: Perspectives in the philosophy of Donald Davidson*, Basil Blackwell, Oxford, pp. 369-386.

KIRK, R. (1973): «Underdetermination of theory and indeterminacy of translation», *Analysis*, 33, pp. 195-201.

— (1982): «Physicalism, identity, and strict implication», *Ratio*, 24, pp. 131-141.

— (1986): «Mental machinery and Goedel», *Synthese*, 66, pp. 437-452.

KITCHER, P. (1984): «In defense of intentional psychology», *Journal of Philosophy*, 81, pp. 89-106.

KOSSLYN, S. M. (1980): *Image and mind*, Harvard University Press, Cambridge, MA.

KRIPKE, S. (1963/1971): «Semantical considerations on modal logic», *Acta Philosophical Fennic*a, 16,pp. 83-94. Reimpreso en L. Linsky(ed.), *Reference and modality*, Oxford University Press, Oxford, pp. 63-72.

— (1971): «Identity and necessity», en M. K. Munitz (ed.), *Identity and individuation*, New York University Press, New York, pp. 135-164. [Version castellana en L. M. Valdés Villanueva, 1991.]

— (1972): «Naming and necessity», en D. Davidson y G. Harman (eds.); *Semantics of natural languages*, Reidel, Dordrecht, pp. 253-355.

LAKOFF, G. (1987): *Women, fire, and dangerous things. What categories reveal about the mind*, University of Chicago Press, Chicago.

LEWIS, D. K. (1966/1971): «An argument for the identity theory», *The Journal of Philosophy*, 63,

pp. 17-25. Reimpreso en D. M. Rosenthal, 1971.

— (1968): «Counterpart theory and quantified modal logic», *Journal of Philosophy*, 65, pp. 113-126.

— (1969/1980): «Review of art, mind, and religion», *Journal of Philosophy*, 66, pp. 23-25. La parte sobre H. Putnam reimpresa en N. Block, 1980a, pp. 232-233.

— (1972/1980): «Psychophysical and theoretical identifications», *Australasian Journal of Philosophy*, 50, pp. 249-258.

— (1980): «Mad pain and martian pain», en *Block*, 1980a, pp. 216-222.

— (1983a): «Postscript to "Mad pain and Martian pain"», *Philosophical papers of David Lewis*, Oxford, New York, vol. 1.

— (1983b): «Individuación by acquaintance and by stipulation», *The Philosophical Review*, 92, pp. 3-32.

LEWONTIN, R. C. (1978): «Adaptation», *Scientific American*, 239, pp. 212-230.

LINSKY, L. (1967): *Referring*, Routledge and Kegan Paul, London.

— (1977): *Names and descriptions*, University of Chicago Press, Chicago.

LOCKE, J. (1959): *An essay concerning human understanding*, Dover, New York. Originalmente publicado en 1690. [Versión castellana de L. Rodríguez Aranda, Ensayo sobre el entendimiento humano, Aguilar, Madrid, 1987.]

LUCAS, J. R. (1961/1964): «Minds, machines, and Gödel», *Philosophy*, 36, pp. 112-127. Reimpreso en J. R. Anderson, 1964, pp. 43-59.

LYCAN, W. G. (1969): «On "intentionality" and the psychological», *American Philosophical Quarterly*, 6, pp. 305-311.

— (1972): «Materialism and Leibniz' Law», *Monist*, 56, pp. 276-287.

— (1974): «Mental states and Putnam's functionalist hypothesis», *Australasian Journal of Philosophy*, 52, pp. 48-62.

— (1981a): «Form, function, and feel», *Journal of Philosophy*, 78, pp. 24-49.

— (1981b): «Psychological laws», *Philosophical Topics*, 12, pp. 9-38.

— (1981c): «Toward a homuncular theory of believing», *Cognition and Brain Theory*, 4, pp. 139-159.

— (1984): *Logical form in natural language*, MIT Press/Bradford Books, Cambridge, MA.

— (1987): *Consciousness*, MIT Press/Bradford Books, Cambridge, MA.

MALCOLM, N. (1984): «Consciousness and causality», en D. M. Armstrong y N. Malcolm (eds.), *Consciousness and causality: A debate on the nature of mind*, Basil Blackwell, Oxford, pp. 3-101.

MALONEY, J. C. (1984): «The mundane mental language: How to do works with things», *Synthese*, 59, pp. 251-294.

— (1985a): «Methodological solipsism reconsidered as a research strategy in cognitive psychology», *Philosophy of Science*, 52, pp. 451-469.

— (1985b): «About being a bat», *Australasian Journal of Philosophy*, 63, pp. 26-49.

— (en preparación): *The mundane matter of the mental langua*ge. MARGOLIS, J. (1976): «The concept of disease», *The Journal of Medicine and Philosophy*, 1, pp. 238-255.

— (1977): «Arguments with intensional and extensional features», *Southern Journal of Philosophy*, 15, pp. 327-339.

— (1978): *Persons and minds: The prospects of nonreductive materialism*, Reidel, Dordrecht.

MAXWELL, G. (1978): «Rigid designators and mind-brain identity», en C. W. Savage (ed.), *Minnesota studies in the philosophy of science*, University of Minnesota Press, Minneapolis, vol. 9, pp. 365-403.

MAYR, E. (1974): «Behavioral programs and evolutionary strategies», *American Scientist*, 62, pp. 650-659.

MCCARTHY, J., y HAYES, P. (1969): «Some philosophical problems from the perspective of artificial intelligence», en B. Meltzer y D. Michie (eds.), *Machine intelligence*, Elsevier, New York, vol. 4, pp. 463-502.

MCCAULEY, R. N. (1986): «Intertheoretic relations and the future of psychology», *Philosophy of Science*, 53, pp. 179-199.

— (1987a): «The role of cognitive explanations in psychology», *Behaviorism*, 15, pp. 27-40.

— (1987b): «The not so happy story of the marriage of linguistic and psychology, or why linguistics has discouraged psychology's recent advances», *Synthese*, 72, pp. 341-354.

MCCLELLAND, J. L., y RUMELHART, D. E., y el Grupo de Investigación PDP (1986): *Parallel distributed processing. Explorations in the microstructures of cognition. Vol. 2: psychological and biological models*, MIT Press/Bradford Books, Cambridge, MA.

MCCLOSKEY, M. (1983): «Intuitive physics», *Scientific American*, 248(4), pp. 122-130.

MCDOWELL, J. (1980): «Meaning, communication and knowledge», en Z. Van Straaten (ed.), *Philosophical subjects*, Oxford University Press, Oxford, pp. 117-139.

MCKEON, R. (1941): *The basic works of Aristotle*, Random House, New York.

MEINONG, A. (1904/1960): «Über Gegenstandstheorie», en *Untersuchungen zur Gegenstandstheorie and Psychologie*, Leipzig. Reimpreso en R. M. Chisholm (ed.), *Realism and the background of phenomenology*, Free Press, Glencoe, IL.

MENDEL, G. (1865): «Versuche über Pflanzen-hybriden», *Verhandlungen des Naturforschenden Vereines in Brünn*, 4, pp. 3-47.

MILL, J. S. (1846): *A system of logic*, Harper, New York.

MINSKY, M. (1975/1981): *A framework for representing knowledge*, Memo #306, Laboratorio de Inteligencia Artificial del MIT, Cambridge, MA. Parcialmente reimpreso en J. Haugeland, 1981, pp. 95-128.

— (1986): *The society of minds*, Simon and Schuster, New York.

MORTENSEN, C. (1978): «Review of Popper, K. R. and Eccles, J. C.», *The self and its brain. The Australasian Journal of Philosophy*, 56, pp. 264-266.

NAGEL, T. (1974/1980): «What is it like to be a bat?», *The Philosophical Review*, 83, pp. 435-450. Reimpreso en N. Block, 1980a, pp. 159-168.

— (1986): *The view from nowhere*, Oxford University Press, Oxford.

NATSOULAS, T. (1981): «Basic problems of consciousness», *Journal of Personality and Social Psychology*, 41, pp. 132-178.

— (1985): «An introduction to the perceptual kind of conception of direct (reflective) consciousness», *Journal of Mind and Behavior*, 6, pp. 333-356.

NEISSER, U. (1975): *Cognition and reality. Principles and implications of cognitive psychology*.

Freeman, San Francisco. [Versión castellana de M. Ato García, *Procesos cognitivos y realidad*, Marova, Madrid, 1981.]

— (1982): *Memory observed*, Freeman, San Francisco.

PALMER, S., y KIMCHI, R. (1986): «The information processing approach to cognition», en T. Knapp y L. Robertson (eds.), *Approaches to cognition: Contrasts and controversies*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 37-77.

PEIRCE, C. S. (1877/1934): «The fixation of belief», *Popular Science Monthly*, 12, pp. 1-15. Incluido en C. Hartshorne y P. Weiss (eds.), *Collected papers of Charles Sanders Peirce*. Harvard University Press, Cambridge, MA, vol. V, pp. 223-247.

— (1878/1934): «How to make our ideas clear», *Popular Science Monthly*, 12, pp. 286-302. En C. Hartshorne y P. Weiss (eds.), *Collected papers of Charles Sanders Peirce*, Harvard University Press, Cambridge, MA, vol. V, pp. 248-271.

PERRY, J. (1977): «Frege on demonstratives», *Philosophical Review*, 86, pp. 464-497.

— (1977): «The problem of the essential indexical», *Nous*, 13, pp. 3-21.

PIATTELLI-PALMARINI, M. (ed.), (1980): *Language and learning: The debate between Jean Piaget and Noam Chomsky*, Harvard University Pres, Cambridge.

PLACE, U. T. (1956/1970): «Is consciousness a brain process?», *The British Journal of Psychology*, 47, pp. 44-50. Reimpreso en C.V. Borst, 1970, pp. 42-51.

— (1888): «Thirty years on-is consciousness still a brain process?», *Australasian Journal of Philosophy*, 66.

POLTEN, E. P. (1973): *Critique of the psyco-physical identity theory*, Mouton. The Hague.

POLLOCK, J. L. (1982): *Language and thought*, Princeton University Press, Princeton.

POPPER, K. y ECCLES, J. (1977): *The self and its brain*, Springer-Verlag, New York. [Versión castellana de C. Solís Santos, *El yo y su cerebro*, Labor, Barcelona, 1985.]

PRATT, J. B. (1922/1975): *Matter and spirit*, Macmillan, New Yok. Algunas partes se encuentran en Mandelbaum, F. W. Gramlick y A. R. Anderson (eds.), *Philosophical problems*, Macmillan, New York, pp. 263-283.

PUTNAM, H. (1960/1964): «Minds and machines», en S. Hook (eds.), *Dimensions of mind*, New York University Press, New York. Reimpreso en A. R. Anderson, 1964, pp. 72-97. [Versión castellana de P. Navarro, en A. M. Turing, H. Putnam y D. Davidson, *Mentes y máquinas*, Tecnos, Madrid, 1985.]

— (1962): «The analytic and the synthetic», en H. Feigl y G. Maxwell (eds.), *Minnesota studies in the philosophy of science*, University of Minnesota Press, Minneapolis, vol. 3, pp. 350-397.

— (1967/1980): «Psychological predicates», en W. H. Capitán y D. D. Merrril (eds.), *Art, mind, and religion*, University of Pittsburgh Press, Pittsburgh, pp. 37-48. Reimpreso como «The nature of mental states», en N. Block, 1980a, pp. 222-231.

— (1973): «Meaning and reference», *Journal of Philosophy*, 70, pp. 609-711.

— (1975a): «Philosophy and our mental life», en H. Putnam (ed.), *Mind, language, and reality: Philosophical papers of Hilary Putnam*, Cambridge University Press, Cambridge, vol. 2, pp. 291-303.

— (1975b): «The meaning of "meaning"», en H. Putnam (ed.), *Mind, language, and reality: Philosophical papers of Hilary Putnam*, Cambridge University Press, Cambridge, vol. 2, pp. 215-271. [Versión castellana en L. M. Valdés Villanueva, 1991.]

— (1978): *Meaning and the moral sciences*, Routledge and Kegan Paul, London.

— (1981): *Reason, truth, and history*, Cambridge University Press, Cambridge. [Versión castellana de J. M. Esteban Cloquell, Razón, verdad e historia, Tecnos, Madrid, 1988.]

— (1983): *Realism and reason*, Cambridge University Press, Cambridge.

— (1984): «Models and modules», *Cognition*, 17, pp. 253-264.

— (1986): «Meaning holism», en L. E. Hahn y R. A. Schlipp, *The Philosophy of W. V. Quine*, Open Court, La Salle, IL, pp. 405-426.

PYLYSHYN, Z. W. (1979/1981): «Complexity and the study of artificial and human intelligence», en M. Ringle (ed.), *Philosophical perspectives in artificial intelligence*, Humanities Press, Atlantic Highlands NJ, pp. 25-56. Reimpreso en J. Haugeland, 1981, pp. 67-94.

— (1980): «Computation and cognition: Issues in the foundations of cognitive science», *The Behavioral and Brain Sciences*, 3, pp. 111-169.

— (1981): «The imagery debate: Analogue media versus tacit knowledge», *Psychological Review*, 88, pp. 16-45.

— (1984): *Computation and cognition: Towards a foundation for cognitive science*, MIT Press/Bradford Books, Cambridge, MA. [Versión castellana de R. Fernandez Gonzalez, *Computación y conocimiento*, Debate, Madrid, 1988.]

QUINE, W. V. O. (1953/1961a): «Two dogmas of empiricism», *From a logical point of view*, 2.ª ed., Harper and Row, New York, pp. 20-46. [Versión castellana en L. M. Valdés Villanueva, 1991.]

— (1953/1961b): «Reference and modality», en *From a logical point of view*, 2.ª ed., Harper and Row, New York, pp. 139-157. [Version castellana de M. Sacristán, *Desde un punto de vista lógico*, 2.ª ed., Orbis, Barcelona, 1985.]

— (1960): *Word and object*, MIT Press, Cambridge, MA. [Versión castellana, *Palabra y objeto*, Labor, Barcelona, 1968.]

— (1969a): «Existence and quantification», en W. V. O. Quine (ed.), *Ontological relativity and other essays*, Columbia University Press, New York. [Versión castellana de J. LI. Blasco, *La relatividad ontológica y otros ensayos*, Tecnos, Madrid, 1986.]

— (1969b): «Replies», en D. Davidson y J. Hintikka (eds.), *Words and objections: Essays on the work of W. V. Quine*, Reidel, Dordrecht, pp. 292-352.

— (1969c): «Ontological relativity», en W. V. O. Quine (ed.), *Ontological relativity and other essays*, Columbia University Press, New York, pp. 26-68. [Versión castellana de J. LI. Blasco y otros, *La relatividad ontológica y otros ensayos*, Tecnos, Madrid, 1986.]

— (1970): «On the reasons for the indeterminacy of translation», *Journal of Philosophy*, 67, pp. 178-183.

— (1973): *The roots of reference*, Open Court, La Salle, IL. [Versión castellana de M. Sacristán, *Las raíces de la referencia*, Alianza, Madrid, 1988.]

— (1975): «On empirically equivalent systems of the world», *Erkenntnis*, 9, pp. 313-328.

REY, G. (1983): «Concepts and stereotypes», *Cognition*, 15, pp. 237-262.

— (1986): «What's really going on in Searle's "Chinese room'», *Philosophical Studies*, 50, pp. 169-185.

RICHARDSON, R. C. (1979): «Functionalism and reductionism», *Philosophy of Science*, 46, pp.

533-558.

— (1980): «Intentional realism or intentional instrumentalism», *Cognition and Brain Theory*, 3, pp. 125-135.

— (1981): «Internal representation: Prologue to a theory of intentionality», *Philosophical Topics*, 12, pp. 171-211.

— (1982): «The "scandal" of Cartesian interactionism», *Mind*, 91, pp. 20-37.

RISTAU, C. (1983): «Language, cognition, and awareness in animáis?», en J. A. Secaer (ed.), *The role of animals in bio-medical research. Annals of the New York Academy of Sciences*, 406, pp. 170-186.

— (1987): *Intentional behavior by birds?: The case of the «injury-feigning» Plovers*, manuscrito no publicado.

RORTY, R. (1965/1971): «Mind-body identity, privacy, and categories», *The Review of Metaphysics*, 19, pp. 24-54. Reimpreso en D. M. Rosenthal, 1971, pp. 174-199.

— (1970/1971): «In defense of eliminative materialism», *The Review of Metaphysics*, 24, pp. 112-121. Reimpreso en Rosenthal, 1971, pp. 223-231.

— (1979): *Philosophy and the mirror of nature*, Princeton University Press, Princeton. [Versión castellana de J. Fernández Zulaica, *La filosofía y el espejo de la naturaleza*, 2.ª ed., Cátedra, Madrid, 1989.]

ROSCH, E. (1975): «Cognitive representations of semantic categories», *Journal of Experimental Psychology: General*, 104, pp. 192-233.

ROSENBERG, A. (1980): *Sociobiology and the preemption of social science*, The Johns Hopkins University Press, Baltimore.

— (1986): «Intention and action among the macromolecules», en N. Rescher (ed.), *Current issues in teleology*, University Press of America, Lanham, MD, pp. 65-76.

ROSENTHAL, D. M. (ed.), (1971): *Materialism and the mind-body problem*, Prentice-Hall, Englewood Cliffs, NJ.

RUMELHART, D. E. (1984): «The emergence of cognitive phenomena from the sub-symbolic processes», *Proceedings of the Sixth Annual Conference of the Cognitive Science Society*, Boulder, CO, pp. 59-62.

RUMELHART, D. E.; MCCLELLAND, J. L. y el Grupo de Investigación PDP (1986): *Parallel distributed processing. Explorations in the microstructure of cognition. Vol. 1: Foundations*, MIT/Bradford Books, Cambridge, MA.

RUSSELL, B. (1905): «On denoting», *Mind*, 14, pp. 479-493. Reimpreso en R. C. Marsch (ed.), Bertrand Russell, *Logic and knowledge*, Capricorn, New York, pp. 41-56. [Versión castellana de J. Muguerza, *Lógica y conocimiento*, 2.ª ed., Taurus, Madrid, 1981.]

— (1940): *An Inquiry into meaning and truth*, George Allen and Unwin, London. [Version castellana de M. A. Galmarini, *Significado y verdad*, Ariel, Barcelona, 1983.]

RUSSOW, L.-M. (1982): «It's not like that to be a bat», *Behaviorism*, 10, pp. 55-63.

— (1984): «Unlocking the Chinese room». *Nature and System*, 6, pp. 221-227.

RYLE, G. (1949): *The concept of mind*, Bames and Noble, New York. [Versión castellana, *El concepto de lo mental*, Paidós, Buenos Aires, 1967.]

SAVAGE-RUMBAUGH, E. S. (1986): *Ape language: From conditioned response to symbol*, Columbia, New York.

SAVAGE-RUMBAUGH, E. S.; MCDONALD, K.; SEVCIK, R.; HOPKINS, W. y RUPERT, E. (1986): «Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (Pan paniscus)», *Journal of Experimental Psychology: General*, 115, pp. 211-235.

SAYRE, K. M. (1986): «Intentionality and information processing: An alternative model for cognitive science», *Behavioral and Brain Sciences*, 9, pp. 121-166.

SCHANK, R. C. y ABELSON, P. (1977): *Scripts, plants, goals and understanding*, Lawrence Eribaum Associates, Hillsdale, NJ.

SCHNAITTER, R. (1987): «Behaviorism is not cognitive and cognitivism is not behavioral», *Behaviorism*, 15, pp. 1-11.

SEARLE, J. R. (1969): *Speech acts: An essay in the philosophy of language*, Cambridge University Press, Cambridge. [Versión castellana de L. M. Valdés Villanueva, *Actos de habla*, 2.ª ed., Cátedra, Madrid, 1986.]

— (1979): *Expression and meaning: Studies in the theory of speech acts*, Cambridge University Press, Cambridge.

— (1980): «Minds, brains, and programs», *The Behavioral and Brain Sciences*, 3, pp. 417-424. Reimpreso en J. Haugeland, 1981a, pp. 282-306.

— (1981): «Intentionality and method», *The Journal of Philosophy*, 78, pp. 720-733.

— (1984): «Intentionality and its place in nature», *Dialéctica*, 38, pp. 87-99.

SELFRIDGE, O. G. (1955): «Pattem recognition by modern computers», *Proceedings of the Western Joint Computer Conference*, Los Angeles, California.

SELLARS, W. F. (1963a): «Philosophy and the scientific image of man», en W. F. Sellare (ed.), *Science, perception, and reality*, Routledge and Kegan Paul, London, pp. 1-40. [Versión castellana de V. Vázquez de Zavala, *Ciencia, percepción y realidad*, Tecnos, Madrid, 1971.]

— (1963b): «Empiricism and the philosophy of mind», en W. F. Sellars (ed.), *Science perception, and reality*, Routledge and Kegan Paul, London, pp. 253-359. [Versión castellana de V. Vázquez de Zavala, *Ciencia, percepción y realidad*, Tecnos, Madrid, 1971.]

SHAFFER, J. A. (1965): «Recent work on the mind-body problem», *American Philosophical Quarterly*, 2, pp. 81-104.

SHANNON, C, y WEAVER. W. (1949): *The mathematical theory of communication*, University of Illinois Press, Urbana, IL. [Versión castellana de T. Bethencourt Machado, *Teoría matemática de la comunicación*, Forja, Madrid, 1981.]

SHER, G. (1975): «Sentences in the brain». *Philosophy and Phenomenological Research* 36, pp. 94-99.

— (1977): «Kripke, Cartesian intuitions, and materialism», *Canadian Journal of Philosophy*, 7, pp. 227-238.

SHOEMAKER , S. (1975/1980): «Functionalism and qualia», *Philosophical Studies*, 27, pp. 291-315. Reimpreso en N. Block, 1980, pp. 251-267.

— (1981): «Absent qualia are impossible-A reply to Block», *The Philosophical Review*, 90, pp. 581-599.

SHORT, T. (1983): «Teleology in nature», *American Philosophical Quarterly*, 20 pp. 311-320.

SIMON, H. A. (1955/1979): «A behavioral model of rational choice», *Quarterly Journal of Economics*, 69, pp. 99-118. Reimpreso en: *Models of Thought*, Yale University Press, New Haven, pp. 7-19.

— (1969): *The sciences of the artificial*, MIT Press, Cambridge.

SKINNER, B. F. (1945/1984): «The operational analysis of psychological terms», *Psychological Review*, 52, pp. 270-277, 291-294. Reimpreso en *The Behavorial and Brain Sciences*, 7, pp. 547-581.

— (1948): *Waiden two*, Macmillan, New York. [Versión castellana, *Waiden Dos*, Martínez Roca, Barcelona, 1984.]

— (1971): *Beyond freedom and dignity*, Knopf, New York. [Versión castellana de J. J. Coy, *Más allá de la libertad y la dignidad*, Martínez Roca, Barcelona, 1986.]

SMART. J. J. C. (1959/1971): «Sensations and brain processes», *Philosophical Review*, 68, pp. 141-156. Reimpreso en D. M. Rosenthal, 1971, pp. 53-66.

SMITH, E. E., y MEDIN. D. L. (1981): *Categories and concepts*, Harvard University Press, Cambridge, MA.

SMITH, G. E., y KOSSLYN. S. M. (1981): «An information-processing theory of mental imagery: A case study of the new mentalistic psychology», en P. D. Asquith y R. N. Giere (eds.), PSA 1980, *Philosophy of Science Association*, East Lansing, MI vol. 2, pp. 247-266.

STALNAKER , R. (1976): «Propositions», en A. MacKay y D. Merrill (eds.), *Issues in the philosophy of language*, Yale University Press, New Haven, pp. 79-91.

STICH, S. P. (1978): «Autonomous psychology and the belief-desire thesis», *Monist* 61, pp. 573-591.

— (1979): «Berween Chomskian rationalism and Popperian empiricism», *British Journal for the Philosophy of Science*, 30, pp. 329-347.

— (1981): «Dennet on intentional systems». *Philosophical Topics*, 12, pp. 39-62.

— (1983): *From folk psychology to cognitive science*, MIT Press, Cambridge, MA.

SUTTON, W. (1903): «The chromosomes in heredity», *Biological Bulletin*, 4 pp. 231-251.

TARSKI. A. (1944/1952): «The semantic conception of truth». *Philosophy and Phenomenological Research*, 4, pp. 341-375. Reimpreso en L. Linsky (ed.), *Semantics and the philosophy of language*. University of Illinois Press, Urbana, pp. 11-47. [Versión castellana en L. M. Valdés Villanueva, 1991.]

— (1967): «Proof and truth», *Scientific American*, 220, pp. 63-77.

TENNANT, N. (1984): «Intentionality, syntactic structure, and the evolution of language», en C. Hookway (ed.), *Minds, machines, and evolution*, Cambridge: Cambridge University Press, Cambridge, pp. 73-103.

THAGARD, P. (1985): *The emergence of meaning: How to escape Searle's Chinese Room*, manuscrito no publicado.

TURING, A. M. (1937): «On computable numbers with an application to the Entscheidungsproblem», *Proceedings of the London Mathematical Society*, 42, pp. 230-265.

— (1950/1964): «Computing machinery and intelligence», *Mind*, 59, pp. 433-460. Reimpreso en A. R. Anderson, 1964, pp. 4-30.

VALDÉS VILLANUEVA , L. M. (ed.), (1991): *La búsqueda del significado*, Tecnos/ Universidad de Murcia, Madrid.

VANGULICK, R . (1986): *A functionalist theory of self-consciousness: The problem*, manuscrito no publicado.

WIERZBICKA, A. (1987): *«Prototypes save»: On the current uses and abuses of the concept «prototype» in current linguistics, philosophy, and psychology*, manuscrito no publicado.

WILKES, K. V. (1981): «Functionalism, psychology, and philosophy of mind», *Philosophical Topics*, 12, pp. 147-167.

WIMSATT, W. C. (1972): «Teleology and the logical structure of function statements», *Studies in the History and Philosophy of Science*, 3, pp. 1-80.

— (1976): «Reductive explanation: A functional account», en R. S. Cohen, C. A. Hooker, A. C. Michalos y J. Van Evra (eds.), PSA-1974. *Boston Studies in the Philosophy of Science*, Reidel, Dordrecht, vol. 32, pp. 671-710.

— (1981): «Robustness, reliability, and overdetermination», en M. B. Brewe y B. E. Collins (eds.), *Scientific inquiry and the social sciences*, Jossey-Bass, San Francisco, pp. 124-163.

WEMOGRAD, T. (1972): «Understanding natural language», *Cognitive Psychology*, 1, pp. 1-191.

— (1981): «What does it mean to understand language?», en D. Norman (ed.), *Perspectives on cognitive science*, Ablex, Norwood, NJ, pp. 231-263.

WITTGENSTEIN, L. (1953): *Philosophical investigations*, Macmillan, New York. [Versión castellana de A. García Suárez y U. Moulines, *Investigaciones filosóficas*, Crítica, Barcelona, 1988.]

— (1958): *The blue and brown books: Preliminary studies for the «Philosophical investigations»*, Harper and Row, New York. [Versión castellana en Tecnos, Madrid, 1968.]

— (1961): *Tractatus logico-philosophicus*, trad, de D. F. Pears y B. F. McGuiness, Routledge and Kegan Paul, London. Publicado originalmente en 1921. [Versión castellana de J. Muñoz e I. Reguera, *Tractatus logico-philosophicus*, 9.ª ed., Alianza, Madrid, 1989.]

WOODHOUSE, M. (1984): *A preface to philosophy*, Wadsworth, Belmont, CA.

WRIGHT, L. (1976): *Teleological explanations: An etiological analysis of goals and functions*, University of California Press, Berkeley.

NOTAS

CAPÍTULO 1

[1] Dentro de la ciencia cognitiva hay ahora filósofos que se ocupan de investigaciones empíricas desarrollando, muy frecuentemente, simulaciones de inteligencia artificial (IA). Esos filósofos están volviendo a una muy antigua tradición en filosofía, ejemplificada por filósofos como Aristóteles, Descartes y Kant, que llevaron a cabo a la vez investigaciones empíricas y desarrollaron análisis más puramente conceptuales. Tales propósitos híbridos no han sido populares en este siglo hasta la pasada década.

[2] Véase Bechtel (en prensa b) para cuestiones adicionales sobre la naturaleza de la epistemología y la metafísica así como también para una discusión sobre los otros campos principales de la filosofía: lógica y teoría moral.

[3] Para una introducción útil a la metodología filosófica, ver Woodhouse, 1984.

[4] Véase Bechtel (en prensa b) para una introducción básica a la lógica moderna y su relevancia para la ciencia cognitiva.

[5] Para Sócrates esto no era necesariamente un defecto. Aunque no se lograban resultados positivos, él parecía contemplar el descubrimiento de que carecíamos de conocimiento y éramos realmente ignorantes como el paso primero y fundamental hacia la sabiduría.

[6] La diferencia para Platón era solamente que los objetivos se definían por medio de las Ideas abstractas que no eran Formas incorporadas.

[7] A estos factores se hace referencia comúnmente como las *cuatro causas* de Aristóteles. El término *causa* es, en este contexto, más bien desorientador. Es por eso por lo que he hablado de factores que necesitaban tomarse en consideración para explicar el cambio.

[8] Esta concepción newtoniana difería de la concepción cartesiana en algunos aspectos importantes. Por ejemplo, Newton y Locke aceptaron la idea de un espacio vacío y de la acción a distancia. Este último concepto era necesario para dar cuenta de las leyes gravitacionales de Newton. Descartes intentó, por otra parte, explicar la gravitación por medio del contacto directo y la interacción de una serie de corpúsculos: de este modo podía evitarse la acción a distancia.

[9] Los mismos newtonianos estaban inclinados hacia el deísmo, una teología de acuerdo con la que Dios era el creador del mundo pero, una vez creado, lo dejaba que operase sólo de acuerdo con sus propios principios. Muchos cristianos, incluido Berkeley, pensaban que este punto de vista apartaba a Dios demasiado del mundo ordinario.

[10] Así Peirce, rechazó también la afirmación de Kant de que hay un dominio de cosas en sí que está más allá del reino del conocimiento. Conoceremos todo lo que hay que conocer cuando la investigación alcance el punto en el que no hay posibilidad de revisión adicional.

CAPÍTULO 2

[1] El papel de tales problemas en el desarrollo de la filosofía del lenguaje moderno es expuesto

claramente por Russell (1905): «Una teoría lógica puede ser puesta a prueba por su capacidad para habérselas con problemas, y es un plan saludable, al pensar sobre lógica, el almacenar en la mente tantos problemas como sea posible, puesto que sirven para los mismos propósitos que sirven los experimentos en la ciencia física» (p. 47).

[2] Merece la pena señalar, sin embargo, que hay otro modo en el que podríamos identificar el referente de una oración: podríamos considerar que es el hecho o el estado de cosas descrito por la oración.

[3] Sócrates, por ejemplo, mantenía que se podría ganar comprensión del conocimiento o de la justicia solamente descubriendo la propiedad esencial que haría de algo una instancia de conocimiento o justicia. Wittgenstein (1958) responde: «La idea de que para obtener claridad sobre el significado de un término general se tiene que encontrar el elemento común en todas sus aplicaciones, ha puesto grilletes a la investigación filosófica; pues no sólo no ha llevado a ningún resultado, sino que también ha hecho que el filósofo se despreocupe, por irrelevantes, de los casos concretos que sólo podrían haberle ayudado a entender el uso de un término general. Cuando Sócrates plantea la pregunta "¿qué es el conocimiento?", ni siquiera considera como una respuesta preliminar el enumerar casos de conocimiento» (pp. 19-20).

[4] Austin no sólo desarrolló este análisis de los actos de habla, sino que invocó el análisis del uso del lenguaje como una herramienta para resolver problemas filosóficos. Esta herramienta requería en primer lugar reunir el vocabulario y los giros usados para hablar sobre un dominio particular, como la responsabilidad, y a continuación examinar con detalle los matices incluidos en el uso de los términos y giros. Para reunir los términos y giros Austin recomendó técnicas tales como la libre asociación, la lectura de documentos relevantes (p. ej., los hallazgos legales sobre la responsabilidad) y el examen de diccionarios. El segundo paso incluía el construir enunciados que podrían usarse efectivamente en el lenguaje, prestando una atención estrecha a qué términos se usarían en el habla normal y qué modos de decir cosas se preferían a otros. Para Austin esta actividad debía llevarse a cabo anteriormente a cualquier teorización filosófica, puesto que tal teorización podría contaminar la evidencia y destruir la sensibilidad hacia cómo la gente usa efectivamente el lenguaje. El objeto de este ejercicio es descubrir las distinciones sutiles hechas en el lenguaje que pueden ser útiles cuando se empiezan a construir teorías filosóficas que, a la vez. den cuenta de cómo se usan normalmente los términos y los giros del lenguaje y se inspiren en las intuiciones sobre el uso ordinario descubiertas en los primeros pasos. En este último paso el enfoque de Austin se aparta del de Wittgenstcin. Para él, el análisis del uso ordinario del lenguaje es un instrumento para ser usado al resolver problemas filosóficos, no nos lleva a disolver los problemas. Austin ilustró este método en estudios sobre tópicos de importancia filosófica, tales como la naturaleza de la responsabilidad humana (Austin, 1956-1957) y los procesos de percepción (Austin, 1962b).

[5] La conclusión de Quine depende del modo particular en el que él interpreta la tesis de la indeterminación. Aunque Quine argumenta a favor de la tesis de la indeterminación sobre la base de la subdeterminación de las teorías, insiste en que la indeterminación no es simplemente la subdeterminación de las teorías lingüísticas (Quine, 1969b, 1970). Incluso si una teoría científica está subdeterminada, podemos hacer ciertas suposiciones teóricas dentro de nuestra teoría y tratarla como una explicación real del mundo. Pero Quine rechaza la idea de tratar una traducción como una teoría sobre lo que significa un lenguaje y permitir la suposición de proposiciones que den cuenta de una hipotética semejanza de

significado. Esta afirmación se ha demostrado que es muy controvertida. Para discusiones adicionales, véase Kirk (1973), Quine (1975) y Bechtel (1978, 1980).

[6] Davidson está haciendo frente aquí a afirmaciones como la de Kuhn de que en las principales revoluciones de la ciencia la nueva teoría es tan radicalmente inconmensurable con la vieja que las dos teorías no pueden discutirse utilizando el mismo lenguaje. Davidson negaría que pudiésemos contemplar a alguien como manteniendo una teoría si no pudiésemos interpretar su teoría en nuestro lenguaje. Para una discusión adicional de los puntos de vista de Kuhn, ver Bechtel (en prensa b).

[7] La segunda oración es efectivamente ambigua. Usando lo que Russell llamó una «distinción de alcance», podemos diferenciar dos lecturas de la oración. Una lectura dice de la persona que era el Presidente número 37 de los Estados Unidos, que esa persona podría no haber sido Presidente. Según esta lectura, a la que se hace referencia comúnmente como la lectura *de re*, la oración es verdadera. La otra lectura dice que es posible que el enunciado «El Presidente número 37 era Presidente» podría haber sido falsa. De acuerdo con esta lectura, a la que se hace referencia como lectura *de dicto*, la oración modal es falsa, puesto que, quienquiera que fuese, el Presidente número 37 era un Presidente. Para discusión adicional, véase Kaplan (1969), Donnellan (1972) y Linsky (1977).

[8] Los defensores de la teoría causal, tales como Rey (1983), han criticado también las propuestas de psicólogos como Rosch (1975) de caracterizar la referencia de términos para géneros naturales mediante prototipos. Siguiendo a Kripke y Putnam, Rey considera que el referente de un término ha de fijarse objetivamente incluso si los usuarios ordinarios del lenguaje no necesitan basarse en diversas estrategias de identificación para determinar si algo es una instancia de un género. Para los teóricos de esta tradición, las estrategias de identificación son independientes del referente, tal como se fija por la teoría causal.

[9] El problema lo desarrolla también Lewis (1968), que niega que pueda haber identidades a través de mundos, sino solamente contrapartidas en un mundo que se parecen estrechamente a individuos de otro mundo.

CAPÍTULO 3

[1] El término *intencionalidad* es un término técnico extraído de la filosofía medieval, donde era usado para referirse a cosas en la mente u operaciones de la mente. Aunque hay una relación entre este término y el término *intencional*, que se deriva de *intender*, ambos no deberían confundirse. El término debe distinguirse también de otro término técnico filosófico relacionado, *intensión*, que se usa comúnmente para referirse al sentido más bien que al referente o extensión de un término.

[2] Hay dispositivos en algunos sistemas biológicos que dan información *sobre* otros estados del sistema e instrumentos hechos por el hombre, como los termostatos y los gasómetros, que realizan tareas similares. Generalmente se piensa que los instrumentos hechos por los hombres derivan su intencionalidad de sus hacedores, mientras que los marcadores de información biológica se han ignorado ampliamente en las discusiones sobre la intencionalidad. Véanse, sin embargo, las discusiones de Dretske y Dennett en el capítulo 4, que proporcionan una perspectiva para considerar la intencionalidad que se encuentra en los sistemas biológicos.

[3] En el capítulo anterior vimos que Meinong tenía un conjunto adicional de razones para extender el

rango de los objetos más allá de los objetos físicos. Allí él usaba objetos puros o subsistentes para explicar la referencia de términos como *unicornio* en oraciones como «Los unicornios no existen» (ver p. 38).

[4] De hecho Frege admite que el sentido de una expresión sirva como referente en contextos tales como el discurso indirecto (ver p. 39).

[5] Brentano introduce el término *relatiforme* para captar este carácter de los estados intencionales: «En el caso de otras relaciones, el Fundamento así como el Término tiene que ser una cosa efectivamente existente... Si una casa es mayor que otra casa, entonces tanto la segunda casa como la primera tienen que existir y tener un cierto tamaño. Pero esto no sucede con las relaciones psíquicas. Si una persona piensa sobre algo, la persona que piensa tiene que existir pero los objetos de sus pensamientos no necesitan existir. Es más, si la persona que piensa está negando o rechazando algo, y si está en lo correcto al hacerlo así, entonces el objeto de su pensamiento no tiene que existir. Así pues, la persona que piensa es la única cosa que necesita existir si ha de darse una relación psíquica. El Término de esta relación no necesita existir en realidad. Uno puede perfectamente preguntar, por tanto, si nos las estamos habiendo con lo que es realmente una relación. Podría decirse efectivamente que estamos tratando con algo que es en ciertos aspectos similar a una relación y que, por tanto, podríamos describir como algo que es *relatiforme* (*etwas "relativliches"*)» (1874, citado en Chisholm, 1967, p. 146).

[6] Hay un tercer aspecto en el que las oraciones sobre estados mentales se desvían a menudo de los principios de la lógica que resultan adecuados para tratar con la mayor parte de los enunciados sobre el mundo empírico. Los enunciados ordinarios sobre el mundo se conforman al principio de la generalización existencial, que nos permite inferir, de una afirmación tal como «Estoy sentado en una silla», la afirmación «Hay algo en lo que estoy sentado».

Algunos enunciados sobre estados mentales violan este principio. El enunciado «Estoy pensando en un unicornio» podría ser verdadero pero no podríamos inferir «Hay algo en lo que estoy pensando».

[7] Este término puede engendrar confusión puesto que este significado de *intensional* debe distinguirse del uso de la palabra *intensión* para hacer referencia a sentidos fregeanos.

[8] La razón que Chisholm añade a esta última condición es que él reconoce que podemos inventar un lenguaje no intencional para describir los estados mentales. Él justifica la tesis original manteniendo que para explicar el significado de las locuciones en este lenguaje no intencional tendríamos que apoyarnos en un lenguaje intencional.

[9] Chisholm considera su explicación de la intencionalidad como algo que destruye los intentos de analizar lo mental en términos de procesos físicos y, de esta manera, unificar las ciencias de la mente con las ciencias físicas. La razón es que hay diferentes principios lógicos que gobiernan los dos dominios (ver Aquila, 1977; Chisholm, 1967). Para un análisis lingüístico diferente de la intencionalidad, ver Anscombe (1965). Ella identifica las oraciones sobre fenómenos intencionales en términos de sus rasgos gramaticales y argumenta que deberíamos analizar la intencionalidad en términos de rasgos gramaticales, no en términos de diferencias ontológicas.

[10] Muchos filósofos, incluyendo los filósofos del lenguaje ordinario y los nominalistas como Quine (ver pp. 74 ss.), han argumentado en contra del intento de analizar el lenguaje en términos de proposiciones. Sin embargo, ha habido un resurgir del interés en las proposiciones, en gran parte como respuesta a los intentos de Carnap, Kripke y Montague de analizar la semántica de la lógica modal. Para

varios análisis contemporáneos de las proposiciones, véanse Harman (1973, 1977), Donnellan (1974), Kaplan (1978), Perry (1979) y Stalnaker (1976).

[11] Una posición relacionada es defendida por Alexander Rosenberg. Él trata las peculiaridades lógicas de las oraciones intencionales como suficientes para expulsar los fenómenos intencionales de la ciencia. Mantiene que hay un hiato insalvable entre el habla intencional y el análisis científico y atribuye nuestra incapacidad para formular leyes verdaderas en términos intencionales a esas peculiaridades lógicas. Observa, sin embargo, que el lenguaje intencional se usa muy libremente en otros dominios tales como la biología molecular o la sociobiología. Se habla, por ejemplo, de que un enzima reconoce un sustrato. Pero él mantiene que en tales contextos esos términos tienen, al menos implícitamente, definiciones conductistas claras. Él piensa que, una vez que estén disponibles definiciones conductistas similares, la psicología será capaz de avanzar. Por consiguiente, recomienda invocar las armazones tanto de la sociobiología como de la biología molecular para desarrollar explicaciones reales de la conducta humana. Ellas reemplazarán los intentos fallidos de desarrollar explicaciones intencionales (Rosenberg, 1980, 1986).

CAPÍTULO 4

[1] Dreyfus (1932) demuestra que este punto de vista ataca tanto a Fodor como a los computadores. Husserl, discípulo de Brentano, desarrolló el punto de vista de que la actividad mental consiste en una gran variedad de actos mentales realizados sobre formas abstractas que el llamó *noemata*. EI enfoque de Husserl se distingue del computacional moderno por el hecho de que él consideró que los *noemata* eran objetos que uno podía examinar conscientemente en aquello a lo que Husserl se refería como la «reducción fenomenológica», mientras que la teoría computacional moderna no está comprometida con ninguna capacidad de la gente de ser conscientes de los símbolos que existen en sus mentes.

[2] La razón por la que no podemos aprender conceptos simplemente por inducción y tenemos que formar y probar hipótesis es que los conceptos sirven para agrupar objetos en clases y hay un número infinito de maneras de hacer esto (ver Goodman, 1955). Tenemos, por tanto, que especificar en una hipótesis los criterios para pertenecer a la clase (Fodor, 1975, p. 36).

[3] Para un punto de vista radicalmente diferente sobre el lenguaje de computación mental, ver Maloney (1984, en preparación). Para la percepción, Maloney propone dejar que los objetos del mundo sirvan como representaciones. Esto resuelve una parte del problema de la intencionalidad, puesto que las representaciones son ahora autorreferenciales. Pero esto no se presta tan fácilmente a un enfoque mentalista de la psicología como lo hace la explicación de Fodor del lenguaje del pensamiento, puesto que es menos claro cómo somos capaces de realizar computaciones con esos objetos.

[4] Los filósofos usan el término *género natural* para referirse a conjuntos de objetos que figuran en las leyes científicas y que tienen condiciones de definición, como, por ejemplo, el oro definido por su número atómico.

[5] Este argumento parece completamente incorrecto puesto que supone que no podemos empezar a articular leyes hasta que se descubran los géneros naturales. Pero éstos solamente pueden descubrirse mediante la búsqueda de regularidades legaliformes. Hay dificultades para identificar factores del entorno que controlan distintas conductas y cogniciones, pero el propósito es encontrar tales

regularidades que nos permitan seleccionar los géneros naturales. Los investigadores pueden proponer y comprobar tales leyes incluso sabiendo qué investigaciones adicionales en psicología o en otras disciplinas pueden forzar su revisión. Maloney (1985a) argumenta también que, si Fodor está en lo correcto, tenemos que esperar el descubrimiento de los géneros naturales, que cuenta también en contra de su psicología solipsista. El precepto de Fodor de ir en contra de la posibilidad de tales leyes no parece tan compulsivo como su argumentación a favor de la necesidad de una psicología computacional.

[6] Muchos investigadores de IA no apoyarían probablemente algunos rasgos del enfoque de Fodor del lenguaje del pensamiento, tales como la afirmación de que el lenguaje del pensamiento tiene que ser innato. Fodor, sin embargo, mantendría que esto es simplemente una consecuencia lógica del punto de vista de la manipulación de símbolos formales que ellos apoyan.

[7] Ver nota 1.

[8] Ortony (en comunicación personal, mayo de 1987) ha sugerido que puede hacerse frente a esta objeción si distinguimos entre creencias representadas y aquellas que son deducibles a partir de las creencias representadas. La plausibilidad de esta respuesta depende de la de desarrollar un conjunto de axiomas a partir del cual no puedan derivarse creencias como las mencionadas por Dennet y Churchland. No es claro que sea posible desarrollar conjuntos de axiomas consistentes para cada persona que generen justamente el conjunto correcto de oraciones al que asignen su creencia cuando se les pregunte.

[9] Para una discusión adicional de estos y otros argumentos en contra de la Teoría Computacional, véanse Amundson (1987), Bailey (1986), Harman (1978), Haroutunian (1983), Hatfield (1986) y Sher (1975).

[10] Ver también Churchland (1979), donde él sugiere que podríamos pensar en la adscripción de actitudes proposicionales como modificaciones adverbiales del modo en que caracterizamos a la gente. Funcionarían entonces de una manera muy similar a como lo hace «rápidamente» en «x se mueve rápidamente».

[11] Para una discusión de esta clase de modelos cognitivos, ver Rumelhart y McClelland (1986) y McClelland y Rumelhart (1986). Es interesante observar que, al defender esta nueva clase de modelos, Rumelhart (1984) hace observaciones que evocan las de los críticos del enfoque computacional. Él ha comentado la inutilidad de desarrollar constantemente explicaciones de la cognición cada vez más complejas basadas en reglas para tratar con aparentes anomalías en la conducta de los agentes cognitivos reales. Él defiende enfoques tipo PDP puesto que éstos son capaces de explicar dentro de una armazón común tanto las conductas que están de acuerdo con las reglas, como las que las violan. Para una introducción a las cuestiones filosóficas planteadas por los modelos conexionistas, ver Bechtel (en prensa c). Para un conjunto de críticas de los modelos conexionistas como alternativas a los modelos computacionales, ver Fodor y Pylyshyn (1987).

[12] Otro filósofo, Sayre (1986), ha intentado combinar la teoría matemática de la información con una perspectiva evolucionista a fin de dar cuenta de la intencionalidad de la percepción. La principal contribución del sistema perceptivo es, para Sayre, centrarse y rastrear las fuentes específicas de información en el entorno de manera que los estados resultantes del cerebro proporcionen al organismo información *sobre* las partes relevantes del entorno. En la explicación de Sayre figura una perspectiva evolucionista puesto que él ha considerado cómo ha evolucionado en los organismos la capacidad de adquirir información centrada, relevante, sobre sus entornos. Para una discusión, ver los comentarios que

siguen al artículo de Sayre y su respuesta.

[13] El *Gedankenexperiment* de Searle pretende ser una réplica de la estructura del diseño de Schank y Abelson (1977) para comprensión de programas. Schank propuso que nosotros, lo mismo que los programas de computador, podríamos comprender historietas usando «guiones», que son estructuras para representar información en términos de rasgos generales de ciertos tipos de eventos. Lo útil de los guiones es que contienen información por defecto de lo que ocurriría en ciertos géneros de episodios. Podemos usar esa información para complementar lo que se nos dice efectivamente en la historieta. El hecho de que nosotros o un programa usemos guiones al comprender una historieta se supone que explica cómo somos capaces de responder a preguntas sobre una historieta cuando la información jamás ha sido enunciada explícitamente en ella. Por razones de simplicidad he dejado los guiones fuera del *Gedankenexperiment* de Searle.

[14] Ver Hamad, 1987; para otras respuestas al ejemplo de Searle de la habitación china, ver Bynum (1985), Carleton (1984), Rey (1986), Russow (1984), y Thagard (1985).

[15] Dennett ha ofrecido también dos argumentos adicionales a favor de un tratamiento instrumentalista de los estados intencionales; ambos pueden también manejarse con el género de explicación que se ha bosquejado aquí. Un argumento descansa en el hecho de que ningún sistema efectivo es completamente racional, mientras que la postura intencional supone racionalidad completa. Esto, sin embargo, puede responderse tratando nuestras adscripciones intencionales iniciales como idealizaciones, muy semejantes a las leyes de los gases idealizadas usadas en física. En una explicación realista, éstas tendrán que modificarse lo que fuera necesario para describir la vida mental efectiva de una persona de una manera muy semejante a como los psicólogos del razonamiento han propuesto teorías de cómo razonamos para dar cuenta de las desviaciones de la lógica normativa (ver Kahneman, Slovic y Tversky, 1982). Para una respuesta crítica diferente a este argumento de Dennett y a las dudas generales sobre el uso de la racionalidad para fundamentar interpretaciones intencionales, ver Stich (1981) y Dennett (1981b) para una réplica. El otro argumento de Dennett señala que las adscripciones de creencia son algunas veces completamente indefinidas, de modo que podemos describir a dos personas como creyendo la misma cosa (Dennett y un químico creen ambos que la sal es cloruro sódico) incluso si hay grandes diferencias en cómo sus creencias se relacionan con otras creencias (por ejemplo, Dennett no puede usar esta creencia para resolver problemas químicos, mientras que el químico sí puede). El problema, sin embargo, puede tratarse usando el instrumento de los mundos nocionales que revela diferencias en el rango de mundos al que se adapta la gente.

[16] Sayre (1986) ha ofrecido un enfoque alternativo para incorporar una explicación de la intencionalidad dentro de una perspectiva evolucionista (ver nota 12).

[17] Cuando Dennett adoptó una perspectiva evolucionista (ver Dennett, 1983), se comprometió tanto con un enfoque adaptativo de la evolución, como con un punto de vista optimizador de la selección natural. El punto de vista adaptativo mantiene que es apropiado explicar cada rasgo como algo que se selecciona a causa de su contribución a la adaptación del organismo, mientras que la interpretación optimizadora ve la selección natural como algo que produce organismos óptimamente adaptados. Estas interpretaciones encajan con el enfoque de Dennett de la postura intencional, que ve esto como una perspectiva normativa o ideal. Sin embargo, han sido severamente criticadas dentro de la biología evolucionista. Gould y Lewontin (1979; Lewontin, 1978) argumentan en contra del punto de vista

adaptativo haciendo observar que no todos los rasgos de los organismos son producto de la selección natural. Para mostrar que algo es producto de la selección natural, es necesario demostrar de una manera ingenieril cómo la selección ha dado lugar de modo efectivo a ese rasgo. Además, los evolucionistas contemplan generalmente la selección natural como un proceso satisfaciente (*satisficing process*), para usar el término de Simón (1955/1979). La selección promueve cualquier rasgo disponible que contribuya a la adaptación, pero no selecciona solamente el más adaptativo. Al desarrollar una perspectiva realista sobre la postura intencional necesitamos tomar en cuenta esas consideraciones evolucionistas. Esto exige abandonar tanto el punto de vista adaptativo como el optimizador y centrarse en cómo nuestros estados cognitivos nos equipan de modo efectivo para vérnoslas con el entorno y cómo, en ocasiones, nos convierten en mal adaptados.

CAPÍTULO 5

[1] En lo que sigue me centro particularmente en porciones del libro conjunto de Popper y Eccles que fueron escritas por Popper.

[2] Popper niega que la realización de operaciones lógicas por los computadores afecte a sus argumentos manteniendo que, puesto que son producto del diseño humano, «tanto el computador como las leyes de la lógica pertenecen enfáticamente a lo que se llama aquí Mundo 3 (Popper y Eccles, 1977, p. 76). Churchland, sin embargo, muestra cómo esta respuesta falla: «¿Interactúa el computador, que es una máquina física, con el Mundo 3 o no? ¿O quizá Popper quiere decir que los estados funcionales de los computadores no son realmente, después de todo, estados físicos? Su réplica pierde absolutamente de vista el objeto de una teoría funcionalista (ver capítulo 7), que consiste en que los estados mentales son estados descritos a un nivel alto de organización funcional y están implementados en los cerebros. Si un sistema francamente físico como un computador puede seguir reglas y procedimientos: puede conformarse a leyes matemáticas, y puede deducir conclusiones que jamás se hayan deducido antes por hombres o por máquinas, entonces es claro que no es necesario plantear como hipótesis mecanismos no físicos basándose meramente en la fuerza de la capacidad de un sistema para seguir reglas y leyes lógicas» (1986. p. 341).

[3] Ryle ofrece un género diferente de análisis de lo que podría llamarse *ocurrencias* mentales —eventos como experimentar una cierta sensación o pensar un cierto pensamiento—. Él propone tratar los eventos como pensar como algo análogo a eventos como hablar: pensar es hablar con uno mismo.

[4] Una alternativa propuesta comúnmente es que vemos cosas que suceden a otros que son similares a cosas que nos suceden a nosotros, y entonces inferimos que la otra persona está sintiendo lo que nosotros hemos sentido en parecidas circunstancias. Wittgenstein rechazó explícitamente este punto de vista sobre la base de que, incluso si hubiera en nosotros un estado interno, no tendríamos fundamento para reidentificarlo posteriormente como el mismo estado. La razón es que, si el estado no es público. no hay manera de comprobar si estamos reidentificando el mismo estado. Podríamos efectivamente olvidar cómo usamos la palabra antes de identificar otro estado.

[5] Existe un problema adicional consistente en que esta misma acción, pedir permiso para hacer una llamada telefónica, puede ser el resultado de muchos estados mentales diferentes (p. ej., creer que el decano quería que yo adquiriese cierta información). La misma disposición conductista podría ligarse

entonces con un número indefinido de estados mentales.

CAPÍTULO 6

[1] Es útil comparar a los eliminativistas con los dualistas. Ambos critican la Teoría de la Identidad haciendo notar que las cosas que decimos sobre los eventos mentales son radicalmente distintas de las que decimos sobre los eventos cerebrales. Los dualistas apelan en este punto a la Ley de Leibniz para mantener que, por tanto, los eventos mentales no pueden ser idénticos a los eventos cerebrales. Los materialistas eliminativos, por otra parte, ven esas diferencias como algo que muestra que nuestro hablar de lo mental nos compromete a decir cosas que son literalmente falsas y que deberíamos, por consiguiente, abandonar el discurso mental en favor del discurso sobre el cerebro.

[2] La inconveniencia de cesar de hablar sobre sensaciones sería tanta, que sólo un materialista fanático pensaría que vale la pena cesar de referirnos a sensaciones. Si se considera que el Teórico de la Identidad está prediciendo que algún día «sensación», «dolor», «imagen mental» y cosas por el estilo serán eliminadas de nuestro vocabulario, puede decirse, con casi total certeza, que está equivocado. Pero, si simplemente está diciendo que *podríamos* eliminar tales términos con un coste no mayor que una reforma lingüística inconveniente, entonces está enteramente justificado. Y considero que esta última afirmación es todo lo que el materialista tradicional ha deseado siempre. (Rorty, 1965/1971, p. 185.)

[3] Una posición estrechamente relacionada con el Materialismo Eliminativo —la sociobiologia— defiende eliminar el enfoque mentalista en favor de uno extraído a partir de la biología evolucionista (ver Rosenberg, 1980).

[4] De acuerdo con Davidson (1970/1980): El punto es... que, cuando usamos los conceptos de creencia, deseo y todos los demás, tenemos que estar preparados, a medida que se acumula la evidencia, para ajustar nuestra teoría a la luz de consideraciones de convicción argumentativa general: el ideal constitutivo de racionalidad controla parcialmente cada fase de la evolución de lo que tiene que ser una teoría en evolución. Una elección arbitraria de esquema de traducción impediría tal falsificación oportunista de la teoría (p. 98).

[5] Al discutir el dualismo en el capítulo anterior, observé que el dualismo de propiedades resultaría ser en muchos aspectos completamente similar a la Teoría de la Identidad como Instancia. Ahora podemos apreciar la similitud. El dualismo de propiedades mantiene que las propiedades mentales constituyen una clase distinta de propiedades que serán verdaderas de los mismos eventos que las propiedades físicas. De acuerdo con la Teoría de la Identidad como Instancia, los eventos pueden clasificarse como eventos mentales o como eventos físicos, dependiendo de qué propiedades se les atribuyan. Esto nos capacita para contemplar la Teoría de la Identidad como Instancia en tanto que postulando conjuntos distintos de propiedades mentales y físicas que podrían ser instanciadas a la vez en el mismo individuo. Por consiguiente, la Teoría de la Identidad como Instancia es muy parecida al dualismo de propiedades.

CAPÍTULO 7

[1] Los funcionalistas australianos, como Smart y Armstrong, así como algunos norteamericanos, como Lewis y Lycan, no diferencian, sin embargo, sus posiciones de la Teoría de la Identidad como Tipo. La caracterización tópico-neutral de Smart de los estados mentales es, de hecho, una caracterización funcional paradigmática. Estos filósofos mantienen, además, que la correspondencia biunívoca entre los estados mentales y los estados físicos no es una parte crítica de su posición. Lo que es crítico es que la tarea realizada por el estado mental se entienda como algo realizado por algunos estados físicos. De este modo, Lewis (1966/1971 y 1972/1980) ha interpretado la identificación funcional de los estados mentales como algo que proporciona la base para una identificación subsiguiente de qué estados físicos los instancian. Cuando encontramos un estado físico que instancia el papel caracterizado funcionalmente, entonces tenemos establecido que es el mismo que el estado identificado funcionalmente. Lewis (1969/1980) ha defendido, además, que acusar a los Teóricos de la Identidad de mantener un tipo de identificación biunívoca entre los estados mentales y los estados físicos es construir un hombre de paja. Él ha mantenido que el teórico de la Identidad siempre ha reconocido la relatividad del contexto en términos de lo que instancia el estado mental. Los ejemplos usados por los primeros teóricos de la Identidad pueden haber sugerido el compromiso con una relación biunívoca, pero esto refleja simplemente un primer intento de formular el punto de vista materialista. Era de esperar que la explicación se tornase más compleja a medida que madurase la Teoría de la Identidad. Los Churchland han criticado también a los funcionalistas como Fodor que ven el Funcionalismo como algo opuesto a las formas reduccionistas de materialismo. Ver P.M. Churchland (1981b).

[2] Para sacar a la luz el carácter abstracto de este género de teoría, Lewis ha propuesto usar la versión ramificada de la teoría que resulta de reemplazar los términos teóricos de la teoría por variables ligadas por cuantificadores existenciales. El objeto de esto es eliminar la mistificación de los términos teóricos y mostrar que su papel en la teoría es completamente caracterizable en términos de las interacciones descritas por la teoría y no de ninguna propiedad intrínseca.

[3] Lewis ha comparado la manera en que esta teoría se presenta con una historieta que un detective podría contar sobre la muerte de un tal Mr. Body: «X, Y y Z han conspirado para asesinar al Mr. Body. Hace diecisiete años, en las minas de oro de Uganda, X era socio de Body... La semana pasada Y y Z se reunieron en un bar de Reading... El jueves por la noche, a las 11.17, Y fue al ático y colocó una bomba de relojería... Diecisiete minutos más tarde, X se encontró con Z en la sala de billar y le dio una pistola cargada... Justamente cuando la bomba estalló en el ático, X disparó tres tiros hacia el estudio a través de los ventanales franceses» (Lewis. 1972. p. 208).

Lo mismo que podemos seguir esta historieta sin saber quiénes son X, Y y Z, Lewis mantiene que podemos seguir el discurso ordinario sobre las relaciones causales de los eventos mentales sin saber qué procesos neurales originan.

[4] La afirmación crítica aquí es la Tesis de Church de que cualquier procedimiento que sea efectivamente computable puede llevarse a cabo mediante un procedimiento recursivo. Esto es sólo una tesis, no un teorema probado, puesto que la noción que incluye de computación efectiva es intuitiva y no se define formalmente. Cuando esta tesis se combina con la explicación de Turing de la Máquina Universal de Turing, que podría computar todas las funciones recursivas, obtenemos la afirmación de que, si la mente emplea procedimientos efectivos, una Máquina Universal de Turing podría llevar a cabo cualquier tarea que la mente puede llevar a cabo.

[5] El criterio de Pylyshyn es bastante similar al que Fodor (1983. 1985) ha usado para identificar módulos en la mente. Esos módulos son dispositivos especializados para realizar tareas cognitivas particulares. Fodor ha mantenido que esos módulos sólo se encontrarán para procesar *inputs* sensoriales y para producir *outputs* y que la psicología sólo será capaz de explicar operaciones cognitivas realizadas a través de esos módulos especializados, no aquéllas llevadas a cabo por estrategias de razonamiento generales. Putnam (1984) ha argumentado a favor de extender la noción de módulo más ampliamente.

[6] Lycan (1981a) caracteriza este proceso vívidamente utilizando una analogía: «Imagínate que tú eres un analista de costes-beneficios de la *Harvard Business School*, contratado por alguna empresa para elevar sus tambaleantes beneficios. En una visita de inspección te presentan a cada uno de los diversos vicepresidentes que dirigen las principales divisiones de la empresa. Preguntas a uno de los vicepresidentes cómo está organizada su división particular: él te presenta a cada uno de los jefes de su departamento. Uno de los departamentos te interesa, y tú preguntas cómo *ese* departamento realiza corporativamente su trabajo. Este proceso continúa hasta un punto final en el que se te muestra una amplia sala llena de administrativos, cada uno de los cuales no hace nada más que clasificar fichas con un índice numerado dentro de un fichero. "¡Éste es el problema!", exclamas tú. "¡Esta gente tiene que ser reemplazada por máquinas!"» (pp. 28-29n).

[7] Este modo de explicación difiere del modelo filosófico estándar, el modelo deductivo-nomológico. El modelo deductivo-nomológico contempla la explicación como un asunto de subsumir una descripción de un evento bajo un principio general o ley a partir del cual puede derivarse la descripción del evento. Para más cosas sobre este modelo, ver Bechtel (en prensa b).

[8] Hay excepciones, incluyendo la obra de Selfidge (1955) sobre los modelos de pandemónium y las apelaciones de Minsky (1986) a la sociedad de mentes y genios.

[9] Otro intento de mostrar una diferencia entre humanos y computadores se basa en el teorema de Gödel, que mantiene que en cualquier axiomatización consistente de la aritmética habrá teoremas indecidibles que, sin embargo, son verdaderos. Lucas (1961/1964), por ejemplo, argumenta: «El teorema de Gödel se tiene que aplicara las máquinas cibernéticas, puesto que pertenece a la esencia de ser una máquina el que sea una instanciación concreta de un sistema formal. Se sigue que, dada cualquier máquina que sea consecuente y que sea capaz de hacer aritmética simple, hay una fórmula que es incapaz de producir como verdadera —esto es: la fórmula es indemostrable-en-el-sistema—, pero que puede verse que es verdadera. Se sigue de aquí que ninguna máquina puede ser un modelo completo o adecuado de la mente, que las mentes son esencialmente diferentes de las máquinas» (pp. 112-113).

Putnam (1960/1964) mantiene que esta objeción reside en una mala aplicación del teorema de Gödel. El sistema del computador, aunque no podría demostrar directamente la oración indecidible, podría demostrar la oración condicional «si la teoría es consecuente, la oración en cuestión es verdadera». Y, mantiene Putnam, ésta es la situación en que nosotros, los humanos, estamos, de modo que el computador puede hacer tanto como pueden tos humanos. (Para una discusión adicional, véase Kirk, 1986.)

[10] A menudo se acusa a los modelos de IA de los sistemas cognitivos de que carecen de tal perspectiva subjetiva. Nagel plantearía la acusación de que no hay nada que sea ser como una máquina. Gunderson (1970/1971) ha desarrollado, sin embargo, una respuesta interesante a esta afirmación. Él mantiene que, aunque no la reconoceremos, la experiencia subjetiva surgirá si construimos sistemas mecánicos que usen las propiedades causales correctas. Esos sistemas, mantiene él, insistirán en que

tienen experiencias. Nuestra suposición de que somos diferentes se debe a lo que él llama la «asimetría» entre los puntos de vista de primera y tercera persona. Nosotros sólo disfrutamos del punto de vista de la primera persona respecto de nosotros mismos, y así no podemos imaginar cómo otras personas podrían tener tales experiencias.

[11] Armstrong ha comparado la situación a un caso en el que podemos describir la figura contenida en la imagen de un rompecabezas pero no podemos ver la figura nosotros mismos. Podríamos ser capaces de hacer esto porque alguien ha señalado qué líneas constituyen la nariz, el pelo, y así sucesivamente, pero no hemos sido capaces todavía de percibir la *gestalt*. Más tarde podríamos llegar a ver la figura misma, pero no aprenderíamos con ello nada nuevo. Tendríamos simplemente una experiencia nueva.

[12] Block y Fodor mencionan que una respuesta al argumento del espectro invertido es afirmar que, por razones que todavía no conocemos, los espectros invertidos son imposibles. Aunque Block y Fodor no han seguido por esta línea, Hardin (1985,1988) ha ofrecido de hecho evidencia de que simplemente los colores fenoménicos no pueden intercambiarse como propone el argumento del espectro invertido. La razón es que.los colores fenoménicos no son simplemente propiedades, sino que tienen una estructura compleja a la que no se puede dar la vuelta. Por tanto, los fundamentos empíricos solos sugieren que el problema del espectro invertido podría ser de poca importancia.

[13] Block escoge la nación china para esta simulación porque supone que se necesitan aproximadamente mil millones de homúnculos para cubrir todos los cuadrados de una Tabla de Máquina para la simulación de una Máquina de Turing, y, si no, él propone que cada homúnculo pueda manejar algunos cuadrados más en lugar de uno solo. La razón de los mil millones es que ése es aproximadamente el número de neuronas del cerebro. Una Máquina de Turing, sin embargo, es quizás el modo más ineficiente de llevar a cabo cualquier procedimiento, y de este modo, como los Churchland argumentan, mil millones es probable que sean drásticamente muchos menos que el número de homúnculos que se necesitan: «Es demostrable que ningún T_m realizado como se describe en la población de China puede simular posiblemente tus relaciones de *input-output*. No hay chinos suficientes, ni siquiera *remotamente*. De hecho un volumen esférico de espacio centrado sobre el Sol y finalizando en la órbita de Plutón sólidamente relleno de chinos como sardinas en lata (aproximadamente 10^{36} homúnculos) no sería todavía ni siquiera remotamente suficiente... Incluso la más humilde de las criaturas está más allá de tal simulación [usando la nación china]. Un gasteropodo como el caracol de mar (*Aplysia California*) tiene más de 332 células sensoriales distintas, y por eso está fuera del alcance de los métodos en disputa... Dejando completamente aparte la cuestión de los *qualia*, la Máquina de Turing china no podría simular ni siquiera una lombriz de tierra» (Churchland y Churchland. 1981, pp. 134-135).

[14] Los Churchland afirman que aprender cosas sobre nuestro sistema nervioso nos ayudará a diferenciar *qualia* más finamente. Al apelar a esas diferencias neurofisiológicas, los Churchland parecen haberse separado de un enfoque funcionalista. pero esto no es el caso. Ellos mantienen que los procesos neurofisiológicos pueden analizarse funcionalmente, así como que no hay división de principio entre las explicaciones funcionales de la mente y de la neurociencia.

[15] Para otros tratamientos de los *qualia*, ver Malcolm (1984), Armstrong (1984) Malonev (1985b), Heil (1983), Horgan (1982, 1984) y Russow (1982). Un problema estrechamente relacionado es el de la conciencia. La mayor parte de los científicos cognitivos no han intentado explicar la conciencia, puesto que parece ser un fenómeno intratable. Pero algunos filósofos y otros científicos cognitivos han

comenzado a analizar la conciencia funcionalmente. Generalmente, han defendido la estrategia de diferenciar aspectos de conciencia y explicar cada uno independientemente. Ver Dennett (1978c, en prensa), Bechtel y Richardson (1983), Natsoulas (1981,1985). Armstrong (1980) y Bricke (1984) para una discusión adicional.

[16] Para un intento diferente de criticar el Funcionalismo mostrando su similitud con el conductismo filosófico, ver Bealer (1978).

[17] Block, sin embargo, ha argumentado, que, de hecho, los argumentos del espectro invertido tienen aún sentido incluso dados los contenidos de una teoría psicológica, mostrando así que los *qualia* no son partes apropiadas de las teorías psicológicas. La afirmación de Block de que los *qualia* no son parte de las teorías psicológicas parece bastante peculiar, puesto que los argumentos de los *qualia* fueron sus primeros argumentos en contra del Funcionalismo de la Psicología Popular y, con todo, vio que esos defectos nos llevaban al psicofuncionalismo. Si la conclusión es que los *qualia* no son propiedades psicológicas, quizás la simulación de mí mismo por la nación china manifieste propiedades psicológicas y el Funcionalismo de la Psicología Popular sea adecuado.

[18] Wimsatt introdujo efectivamente diversos usos de enunciados funcionales además de la versión explicativa que hemos estado considerando. Uno de ellos es un uso evaluativo, puesto que éste nos permite mirar hacia las fuerzas de selección actuales y no a las que operan históricamente. Wimsatt no diferencia claramente entre estas dos versiones de los enunciados de función, y para él el uso explicativo es primario. No estoy desestimando la importancia del género de explicación evolucionista hacia la que señalan tanto Wimsatt como Wright (ver también Falk, 1981), sino que estoy argumentado que es el análisis funcional el que constituye la armazón teleológica primaria y el que se necesita para introducir una perspectiva teleológica dentro de los análisis psicológicos de los estados mentales.

[19] Este esquema evolucionista no nos compromete con sociobiologfa. Podemos pensar que las estrategias cognitivas evolucionan para cumplir las necesidades de la evolución sin reducirlas a adaptaciones genéticamente codificadas. Para una perspectiva evolucionista sobre los fenómenos culturales no sociobiológica, ver Boyd y Richerson (1985).