

LA MENTE EN EL MUNDO
INTRODUCCIÓN AL PROYECTO DE LA COGNICIÓN CORPORIZADA

ADRIÁN ESPINOSA BARRIOS



Universidad Autónoma de la Ciudad de México

La mente en el mundo

Introducción al proyecto de la cognición corporizada

Universidad Autónoma de la Ciudad de México

Dra. Tania Hogla Rodríguez Mora
Rectora

M. en I. Raúl Amilcar Santos Magaña
Coordinador Académico

Mtra. Marissa Reyes Godínez
Coordinadora de Difusión Cultural y Extensión Universitaria

Equipo de la Biblioteca del Estudiante

Ángeles Godínez Guevara
Responsable

Ana Beatriz Alonso Osorio
Florina Piña Cancino
Miguel Napoléon Estrada Serrano
Sergio Javier Cortés Becerril
Verónica Durán Carmona

La mente en el mundo

Introducción al proyecto de la cognición corporizada

Adrián Espinosa Barrios

FICHA CATALOGRÁFICA E-S/N

Espinosa Barrios, Adrián, autor.

La mente en el mundo : introducción al proyecto de la cognición corporizada / Adrián Espinosa Barrios. — Primera edición. -- México : Universidad Autónoma de la Ciudad de México, 2022

182 páginas ; 21 cm.

Bibliografía: páginas 163-174

ISBN 978-607-8840-10-6

1. Mente y cuerpo. – 2. Filosofía de la mente. – 3. Filosofía y ciencia cognitiva. -- I. título.

LC BD418.5

Dewey 128.2

D.R. © Adrián Espinosa Barrios *La mente en el mundo. Introducción al proyecto de la cognición corporizada*

D.R. © Universidad Autónoma de la Ciudad de México
García Diego 168, col. Doctores,
alc. Cuauhtémoc, c. p. 06720, México, D F

primera edición, 2022

ISBN: 978-607-8840-10-6

Biblioteca del Estudiante:
https://www.uacm.edu.mx/Organizacion/CoordinacionAcademica/Biblioteca_Estudiante

Material educativo universitario de distribución gratuita para estudiantes de la UACM. Prohibida su venta

Hecho e impreso en México

PRESENTACIÓN

Este libro forma parte del acervo de la Biblioteca del Estudiante de la UACM. Su objetivo principal es servir como libro de texto para la materia de *Filosofía de la mente*. Su contenido servirá también como libro de referencia para materias como *Epistemología clásica y no clásica de la ciencia*, así como para *Teoría del conocimiento*. Todas ellas de la carrera de Filosofía e Historia de las Ideas. No obstante, puede servir también como texto de referencia para estudiantes de otras carreras que estén interesados en conocer las líneas generales del trabajo actual en filosofía de la mente. El público interesado en la filosofía actual ha de encontrar aquí un texto accesible y autocontenido que presenta en un lenguaje no muy técnico algunos de los debates contemporáneos tanto en filosofía como en el campo de la neurociencia cognitiva.

La filosofía hoy está más viva que nunca. Sin embargo, buena parte del trabajo que hoy se hace en el área no responde a la idea tradicional de la filosofía como trabajo meramente especulativo ajeno a la realidad. Hoy, la filosofía interactúa de diversos modos con otras disciplinas integrándose a programas de investigación de amplio alcance. Este libro

pretende abrir una pequeña ventana a esa forma de trabajo científicamente informado en lo que respecta a la filosofía de la mente. Así, cualquier persona interesada en entablar puentes de comunicación entre disciplinas sobre el área en cuestión puede encontrar aquí una lectura de su interés.

En buena medida, el libro se ha configurado con base en apuntes de clase. Para ello me ha sido útil la retroalimentación de varias personas que han tenido a bien leer fragmentos del texto para darme sus comentarios. Estoy en deuda con todos ellos. Por supuesto, ninguno es responsable de las omisiones y erratas finales del libro. En primer lugar, agradezco a la Dra. Leticia Romero Chumacero sin cuya sugerencia y apoyo este libro no hubiera sido posible. Especialmente agradezco a Norma Barrios Espino, mi madre, por su lectura dedicada y médicamente informada. Su interés y cariño al revisar lo escrito no tienen forma suficiente de agradecerse. A Andrea Guerrero por ayudarme a entender y desarrollar las ideas en torno a la filosofía de la psicopatología, particularmente en lo concerniente a los delirios y al síndrome de Cotard. A Diana Martínez Lechuga por su dedicado e incansable trabajo tanto en la lectura y corrección del texto, así como en la revisión del aparato crítico. A Joaquín Morales Reyes por su ayuda bibliográfica y en la implementación pedagógica del texto. Ambos prestaron su apoyo como parte del proyecto de servicio social Programa de Estudios Sociales sobre Ciencia y Tecnología con la clave UACM/ss/16-16/099/INT. Agradezco también a los estudiantes de la materia de Filosofía de la mente a lo largo de los últimos años, particularmente a los del curso 2019-II en el plantel Cuauhtepic por haber colaborado con discusiones y con la lectura del primer borrador del libro.

Finalmente, agradezco a la Universidad Autónoma de la Ciudad de México, en particular a la Coordinación Académica, por las facilidades para la elaboración y publicación de este libro.

ADRIÁN ESPINOSA BARRIOS
Ciudad de México, noviembre de 2021

INTRODUCCIÓN

Somos una mente. Pero también somos un cuerpo. En tanto individuos, somos una unidad *psicosomática*. Somos capaces de identificar en forma analítica dos aspectos de esa unidad: mente y cuerpo, pero eso no significa que se trate de entidades aisladas e individuales. Por lo que sabemos, tanto por la ciencia como por nuestra experiencia cotidiana, la mente no es «algo» aislado ni subsistente por sí mismo, sino que se encuentra anclada al mundo físico. Es decir que la mente *está* en el mundo; la mente *es* mundo. Hoy, la biología nos permite recrear el proceso por el cual la mente, nuestra mente, apareció en el mundo. A lo largo de millones de años las condiciones ecológicas de nuestro planeta impusieron presiones evolutivas a la vida que permitieron el desarrollo de especies con un sistema nervioso. La lucha por lograr la sobrevivencia en contextos con recursos limitados, en conjunción con la variación genética ciega y azarosa, permitió que esa versión primitiva del sistema nervioso se especializara y complejizara. Con el tiempo, surgieron los primeros cerebros.

Las especies cerebralizadas encontraron ventajas reproductivas que permitieron la herencia de su estructura genética. Con ello, el

sistema nervioso encontró múltiples formas de especialización que permitieron que las especies presentaran adaptaciones cada vez más eficaces a sus nichos ecológicos. Los homínidos aprovecharon en formas inéditas y creativas estas ventajas para construir la estructura estable de sobrevivencia que hoy llamamos cultura. Una de estas especies, la *sapiens*, desarrolló su cerebro aceleradamente con una adición evolutiva de tremenda eficacia: el *neocortex*. La actividad de ese cerebro le permitió al *sapiens* identificar sus emociones y pensamientos. La variada vida mental de la especie es un producto de la compleja actividad del sistema nervioso que posee y de su interacción con el medio en el que se encuentra. Y es esa misma actividad la que le permite al *sapiens* tener la facultad de pensamiento representativo y abstracto que se requiere para preguntarse por la esencia y función de la propia mente.

Con sus 86 mil millones de unidades y sus cientos de miles de millones de conexiones, el cerebro humano es el órgano conocido más complejo generado por la evolución; y la mente, responsable de toda la vida cultural, emotiva y racional humana, es su producto emergente más asombroso. Es como si, repentinamente, nos percatáramos que siempre estuvimos en posesión y uso de una compleja y versátil herramienta que hasta ahora reconocemos y estudiamos con pasmosa fascinación. Es por eso que, en décadas recientes, han cobrado importancia las disciplinas que tienen al complejo *mente-cerebro* como su objeto de estudio.

En julio de 1990 el entonces presidente de los Estados Unidos, George W. Bush, proclamaba a la década que iniciaba como la «década del cerebro». Con ello, se garantizaba el apoyo económico para impulsar el trabajo de ciencias como la psicología cognitiva y las neurociencias en una forma nunca antes vista. Pero muchas otras disciplinas científicas también se vieron beneficiadas por este impulso, haciendo progresar nuestro conocimiento sobre el cerebro de

muy variadas maneras: produciendo modelos matemáticos para la comprensión de las funciones del sistema nervioso, desarrollando métodos y técnicas de observación de las funciones mentales sin necesidad de realizar trepanaciones craneales o diseñando tecnología para la mejoría o cura de diversas deficiencias neurológicas.

A pesar de que la tarea de comprender el cerebro en su totalidad es titánica, nuestro conocimiento científico sobre su estructura y función avanza a pasos agigantados. Sin embargo, aún imaginando que tal conocimiento sea completo algún día, parece que todavía quedarían pendientes de responder las preguntas relacionadas con la mente: ¿Qué es la mente?, ¿cómo se relaciona con el cerebro?, ¿cómo es posible que los sucesos electroquímicos del sistema nervioso produzcan nuestra compleja vida mental?, ¿es posible conciliar en una sola teoría ambas descripciones (la científica y la fenoménica)?, esa posible teoría, ¿sería científica o filosófica?, ¿o tendríamos que crear una teoría híbrida (transdisciplinaria) que involucre recursos, léxicos y metodologías tanto científicas como filosóficas?, ¿cómo se podría hacer eso? La respuesta a estas preguntas no proviene propiamente de la ciencia empírica, sino de la filosofía.

El hecho obvio de que se requiere de una mente para preguntar sobre ella misma implica una reflexividad que sólo se encuentra en la filosofía. Así, la reflexión filosófica había sido la única vía de investigación hasta el surgimiento de las ciencias empíricas. Gracias a esa reflexión se pueden abordar los aspectos normativos del conocimiento y las descripciones de las estructuras de nuestra experiencia en primera persona. Gracias a ella, también podemos explorar los límites y alcances de nuestras facultades y conceptos mentales tales como: percepción, razonamiento, conciencia, decisión, intencionalidad, etcétera. En los diferentes sistemas filosóficos, desde los clásicos griegos, hasta los modernos y contemporáneos podemos encontrar teorías que intentan dar respuestas reflexivas a estos temas.

Sin embargo, con frecuencia las discusiones filosóficas se entrampan en especulaciones teóricas limitadas por la validez de los argumentos más que por las condiciones reales de los objetos de los que se trata. Y eso no es propiamente incorrecto; la argumentación filosófica se encuentra en el ámbito de lo lógicamente posible, es decir, la filosofía puede establecer el espectro del discurso válido; pero, la identificación de las características de la realidad empírica en que nos encontramos es el ámbito de la ciencia empírica. Una forma abreviada de decirlo es ésta: la filosofía se encarga de la validez del discurso significativo (tanto en el mundo real como en todo mundo posible), mientras que la ciencia se encarga de la confirmación de la verdad o falsedad de las afirmaciones sobre el mundo real. Es por eso que, mientras que la filosofía realiza experimentos mentales en los que la única regla es evitar las contradicciones lógicas, la ciencia recurre a la experimentación empírica.¹

Si esto es cierto, sería de esperar que la respuesta a una comprensión completa de la mente provenga, como hemos mencionado, de una interacción fecunda entre una reflexión filosófica y una exploración científica. Sin embargo, esto no ha sido del todo obvio desde un inicio. Durante varias décadas, ya muy entrado el siglo xx, el problema permaneció aislado en ambos frentes. Los filósofos intentaban proporcionar una teoría reductiva que solucionara de una vez por todas, el problema de la relación entre la mente y el cerebro, mientras que desde el terreno científico se avanzaba en la identificación experimental de la comprensión psicofísica. Pero la intersección sería

¹ Varios de los temas que trataremos en este libro giran en torno a experimentos mentales que se han vuelto famosos: los zombies filosóficos, el cuarto chino, el argumento del conocimiento o el espectro invertido son sólo algunos cuantos ejemplos. Aprender a diseñar experimentos de este tipo, así como su defensa o refutación requieren de habilidades lógico-filosóficas que este texto pretende ayudar a desarrollar.

inevitable. Eventualmente, especialistas de ambos grupos y de muy diversas perspectivas teóricas voltearon hacia otras disciplinas en busca de recursos teóricos y metodológicos que les permitieran abordar un problema que a todas luces rebasaba cualquier intento aislado de comprensión. Y es que actualmente hay una creciente tendencia en reconocer que el binomio de la mente-cerebro es un problema complejo del que no se pueden esperar soluciones simples.

Así, después del surgimiento del área conocida como ciencias cognitivas, y a partir de nuevas formas de comprensión filosófica de la mente enriquecidas por los temas de la conciencia y de la cualidad subjetiva de la experiencia, se desarrollaron programas de investigación que asumen que la comprensión de la mente debería de provenir de aspectos tanto reflexivos como descriptivos, al tiempo que se había de comprender a la mente en su acción ecológica a través del cuerpo. Se trata de proyectos que tienen su base conceptual en teorías como la de la psicología ecológica, la cognición corporizada, la mente situada y la mente extendida.

Otro aspecto relevante en la investigación actual de la mente proviene del área de la psicopatología. A diferencia de lo que se había hecho desde la filosofía durante siglos, hoy se piensa que una fuente de especial relevancia para el conocimiento del funcionamiento de la mente no está nada más en su normalidad sino, justamente, ahí en donde deja de funcionar como debiera. Las formas en que nuestra mente puede fallar dicen mucho sobre sus límites y condiciones. Es una fuente rica de información que la filosofía no puede pasar por alto.

Actualmente, las interacciones son múltiples. Los problemas que se discuten dentro del área de la filosofía de la mente y las neurociencias cognitivas se han diversificado y se alimentan constantemente de los hallazgos científicos más recientes. Sin duda, podemos afirmar que se trata de una de las áreas de investigación más relevantes y ac-

tivas de los años recientes. Y la perspectiva en el futuro inmediato es muy prometedora. Aquí abordaremos en forma introductoria sólo una perspectiva: la filosófica, en particular, los proyectos que toman en cuenta la corporización de las funciones mentales, sus patologías y su acción en el entorno. Para ello, habremos primero de exponer en qué consiste la perspectiva tradicional, no corporizada, para después dar paso a los proyectos que asumen a la mente como una entidad constitutivamente mundanizada.

En su conjunto, el libro tiene como objetivo presentar a los lectores la investigación en torno a la mente como el estudio transdisciplinario de un sistema complejo, que sea capaz de superar la imagen de que lo propiamente «mental» radica únicamente en las funciones cognitivas, poniendo así énfasis en las interacciones medioambientales y sociales a partir de la comprensión de los estados afectivos, volitivos y emocionales que son parte constitutiva de nuestra vida consciente.

En el primer capítulo estudiaremos brevemente las teorías filosóficas que pretenden explicar la relación mente-cuerpo en forma reductiva: fisicismo, funcionalismo, conductismo. Posteriormente, se presenta la Teoría Representacional Computacional de la Mente (TRCM) que dominó la investigación durante varias décadas y que se verá cuestionada por la perspectiva corporizada. El segundo capítulo trata, precisamente, de las teorías que forman parte del espectro de la corporización. Se hace una exposición general de los principios teóricos comunes a estas teorías y se presentan a los principales autores y sus ideas. Finalmente, el capítulo tercero presenta la interacción entre la filosofía y la psicopatología haciendo énfasis en algunas de las deficiencias que resultan de mayor relevancia filosófica.

I. LA MENTE AISLADA

Durante la segunda mitad del siglo XIX se ve puesta a prueba la confianza depositada en la ciencia durante los últimos doscientos años. La certeza que habíamos ganado en Occidente como producto de la efectiva aplicación de los principios sentados por genios como Euclides, Galileo, Kepler y Newton, se tambalea con el hallazgo de fenómenos que se resisten a ser explicados por las leyes que hasta entonces dábamos por verdaderas. La geometría euclidiana cede su lugar de privilegio en la descripción del mundo físico al ver refutado uno de sus axiomas; la mecánica clásica parece funcionar bien en el terreno mesoscópico, pero se encuentra con dificultades en contextos que involucran distancias astronómicas, velocidades cercanas a la de la luz o espacios subatómicos.¹

¹ Para una revisión novedosa y polémica de la Revolución científica, se puede consultar el texto de David Wooton (2017) *La invención de la ciencia*. A diferencia de otros autores, Wooton propone que el periodo comprendido entre Tycho Brahe y Newton presenta una actitud hacia la naturaleza totalmente inédita en el periodo anterior, por lo que no se puede afirmar que la ciencia moderna sea el producto de un proceso histórico gradual.

Como producto de estas crisis, los propios científicos se ven obligados a reflexionar sobre las condiciones epistémicas y ontológicas de sus propias disciplinas con la finalidad de regresar a los principios básicos a partir de los cuales se pueda volver a construir el edificio sólido del conocimiento. El planteamiento de estas preguntas, tanto como sus posibles respuestas, trasciende en forma esencial los alcances de la ciencia. Se trata de reflexiones de carácter filosófico. Así, son los propios científicos (físicos y matemáticos, principalmente) los primeros en desarrollar lo que hoy conocemos como la filosofía de la ciencia contemporánea.² Las preguntas propias de esta naciente disciplina intentaban encontrar explicaciones para diversos temas, como: las razones detrás del cambio entre teorías, la forma en que los enunciados teóricos se relacionan entre ellos y con la observación, la vía por la que se generan y validan los conceptos científicos, la manera en que se relacionan jerárquicamente las ciencias entre sí y cómo pueden éstas ciencias ser axiomatizadas, es decir, cómo es posible excluir de ellas la posibilidad de la contradicción (creadora de crisis) de una vez por todas.

Esta empresa se abordó con un ánimo de esperanza y trabajo colectivo, principalmente, por el grupo conocido como Círculo de Viena. Este ánimo puede explicar el hecho de que, a pesar de las crisis epistémicas al interior de las diversas ciencias, siga siendo su método reductivo el modelo a través del cual se intentan encontrar explicaciones científicas sobre la esencia y función de la mente.

² Sobre el desarrollo de la filosofía de la ciencia contemporánea, se puede consultar el excelente y breve libro de Ulises Moulines (2011), *El desarrollo moderno de la filosofía de la ciencia*.

Para principios del siglo xx, la filosofía revive el problema clásico planteado por Descartes 300 años antes sobre la forma en que se relacionan la mente y el cuerpo. La cuestión puede plantearse de la siguiente manera: toda vez que entendemos que el ser humano parece ser una unidad funcional compuesta tanto de mente como de cuerpo, ¿cómo es que ambas entidades se relacionan entre sí dado que parecen tener propiedades contradictorias?³ La forma en que Descartes planteaba el problema, tanto como su presunta solución, partían de una *concepción dualista*. Es decir, para el filósofo francés se partía del hecho de que mente y cuerpo eran de inicio dos cosas distintas, por lo que aquello que habría que explicar es la forma en que dichas sustancias se relacionan entre sí. Que ambas sustancias se comunican es una evidencia inmediata para toda persona: la intención de mover mi brazo derecho que me formo en la mente se traduce sin dificultad en movimiento; de la misma manera, el golpe accidental en el dedo meñique de mi pie izquierdo contra la pata de la cama se comunica de inmediato en mi mente en forma de dolor. Descartes cree encontrar la razón de este lazo *psicofísico* en una estructura cerebral determinada: *la glándula pineal*.⁴ En esta estructura, afirmaba Descartes, debían existir unos «animales espirituales» que tuvieran cualidades tanto extensas como mentales. Como se puede

³ Al respecto, véase el libro de Descartes *Meditaciones metafísicas*, particularmente, las meditaciones II y VI.

⁴ Hoy en día sabemos que la glándula pineal, o epífisis cerebral, no tiene ese papel metafísico asignado por Descartes, sino uno más modesto, pero vital: entre otras cosas, ayuda a regular los patrones de sueño y vigilia a partir de la producción de melatonina.

observar, esta respuesta, lejos de resolver el problema, lo multiplica al infinito en escalas cada vez más pequeñas.

Los filósofos posteriores a Descartes retomaron el reto de la comunicación de las sustancias. Malebranche, quien recibió la iluminación filosófica al hojear en una librería parisina el texto de Descartes, «resolvió» el problema afirmando que las sustancias no se comunicaban en forma directa, sino a través de otra sustancia: Dios. De tal modo, mi intención de mover mi brazo derecho es conocida por Dios, quien, en su infinita bondad, permite que dicha acción suceda; de igual modo, el dolor en mi mente es insertado por Dios al percartarse que me golpeo el dedo meñique del pie contra la cama. A esta doctrina se le conoce como *ocasionalismo*.

Leibniz también apeló a Dios, pero no como una forma de mensajero entre las sustancias, sino como organizador perfecto y racional de las todas las interacciones posibles entre mente y cuerpo por toda la historia; según su teoría de la *armonía preestablecida*, las sustancias en realidad no se comunican, sino que se encuentran en una armonía sincrónica perfecta desde el inicio de los tiempos.

Para Spinoza, en cambio, la solución no era ontológica, sino semántica: no existen tres sustancias, sino una sola (Dios) con manifestaciones tanto físicas como mentales. Esta es la teoría del *doble aspecto*.

Más allá de las diferencias en la forma de resolver el problema cartesiano, la mayoría de las teorías racionalistas del siglo xvii asumieron una forma de dualismo. Pero esto cambiaría con la generación de nuevo conocimiento proporcionado por las ciencias empíricas. En el siglo xx, la forma de comprender el problema daría un giro importante. Los éxitos de la matematización de la naturaleza, por un lado, y

el positivismo como doctrina epistemológica por el otro, sentaron las bases para una forma de comprensión del mundo de carácter *reductivo*: lo complejo se explica por lo simple. De tal modo, toda teoría, problema o disciplina compleja, puede ser «reducida» a un lenguaje más simple (aunque más extenso) sin pérdida de significado. Esto quiere decir que los conceptos abstractos no designan propiamente objetos, sino que son abreviaturas para la composición de elementos más simples cuya existencia puede ser constatada por la experiencia directa. Por lo tanto, los valores que se privilegian son los de la unidad, la simplicidad y la reducción.⁵

Surge entonces un espectro de teorías sobre la mente, las cuales pretenden explicar su complejidad apelando a equivalencias lógicas con lenguajes teóricos más simples (de nivel de abstracción más bajo pero que designan conceptos y leyes más fundamentales). Tenemos al fisicismo, el conductismo y el funcionalismo, como principales exponentes. Para entonces, los dualismos de carácter cartesiano entran en franco descrédito.

Al mismo tiempo, las nuevas teorías de la información, por un lado, y el diseño y construcción de nuevos dispositivos computacionales por el otro, darían lugar a dos ideas poderosas sobre el funcionamiento de la mente: la *computación* y la *representación*. Bajo estas ideas, los estados mentales no son otra cosa más que representaciones del mundo con un contenido que puede ser manipulado computacionalmente en forma simbólica. De tal modo, es posible, en principio, entender la estructura y funcionamiento de la mente si en-

⁵ Existen diferentes nociones de reducción en la filosofía de la mente. Al respecto, véase el siguiente apartado.

tendemos la forma en que las representaciones adquieren su contenido y la manera en que operan las relaciones lógicas entre ellas.

En este primer capítulo haremos una rápida revisión de esas tres ideas que delinearon la forma de entender la relación mente-cerebro desde la filosofía: la *reducción*, la *computación* y la *representación*. En conjunto, estas ideas dieron forma a lo que podemos llamar el esquema *cognitivo-computacional*, o *Teoría Computacional Representacional de la Mente* (TCRM), que dominó la filosofía de la mente hacia la mitad del siglo XX y cuyas consecuencias se han filtrado a nuestra comprensión cotidiana de la inteligencia, la conciencia y la vida mental en general.

1. Filosofía de la mente: explicando por reducción

Existen múltiples nociones de reducción. En buena medida, se puede afirmar que la filosofía de la ciencia tuvo un carácter reduccionista en su primera etapa. Cada teoría que pretendía explicar la cualidad esencial de la ciencia asumía que dicha cualidad era necesaria y suficiente para la comprensión total del conocimiento científico. Es por eso que las diferencias entre esas teorías estaban determinadas por la forma en que se comprendía tal reducción.

Como veremos, esa cualidad reductiva, junto con la variedad de significados propia del término, está presente en las teorías que pretenden aclarar la relación mente-cerebro. A pesar de sus diferencias específicas, toda teoría de la reducción tiende a suscribir alguna forma de cuando menos una de las siguientes afirmaciones:

1. Existe un objeto, concepto, lenguaje, ámbito, ciencia, etcétera, que puede ser reducido (explicado, analizado) por otro de un nivel de generalidad más básico sin pérdida de significado.
2. El elemento (lenguaje, ciencia, etcétera) *reductor* tiene una cualidad composicional de modo que puede armar cadenas de expresiones significativas para *todo* término del lenguaje *reducido*.
3. Se afirma que existe una cualidad *necesaria* (de la que no se puede prescindir) y *suficiente* (no hace falta nada más para identificar plenamente al objeto a reducir).
4. La estructura de la reducción puede tomar la forma lógica de una implicación $(\forall x) \Phi x \rightarrow (\forall x) \Psi x$ (en caso de que Φ sea un lenguaje o conjunto más amplio y básico de términos u objetos que el representado por Ψ) o de una equivalencia: $(\forall x) \Phi x \leftrightarrow (\forall x) \Psi x$ (en caso de que entre Φ e Ψ haya identidad).
5. La reducción puede operar en un nivel ontológico al hacer una afirmación del tipo: todos los objetos del conjunto x no son más que objetos del conjunto y .
6. La reducción puede operar a un nivel epistemológico: todo el significado de las expresiones del lenguaje x en realidad queda incluido en las expresiones del lenguaje y .

Fisicismo:⁶ la mente es cerebro

Las teorías dualistas fueron la primera respuesta filosófica al problema de la relación mente-cuerpo. La misma presentación del problema sugería ya una división entre dos elementos cuya relación había que explicar o justificar. La postura contraria al dualismo sería un *monismo*: la idea de que sólo existe un elemento de la presunta dualidad y, por tanto, no hay una relación que buscar ni que explicar. Si bien es cierto que existen ejemplos de teorías monistas idealistas (el de Berkeley es el caso más representativo), la mayoría de los monismos son de carácter fisicista. En realidad, el término *fisicismo* es una etiqueta amplia que agrupa diversas posturas reductivas de carácter materialista. No todas estas posturas coinciden en las mismas afirmaciones y ni siquiera pueden ser todas ellas agrupadas como teorías de la relación mente-cuerpo. Algunas de ellas se identifican más con otras posturas reductivas como con los conductismos o los funcionalismos, mientras que otras resultan tan radicales que sus afirmaciones son poco compartidas por otras teorías fisicistas.

Utilizaremos aquí el término fisicismo para agrupar estas teorías bajo una misma familia en la que se comparte un núcleo central de afirmaciones en común, pero con el cuidado de hacer las distinciones necesarias a lo largo de este capítulo. Así, además del fisicismo propiamente dicho, entenderemos

⁶ Se utilizará el término «fisicismo» para traducir el inglés «physicalism». Es traducción común en español el término «fiscalismo», sin embargo, puesto que no hay en nuestro idioma el adjetivo «fiscal», y sí por el contrario el adjetivo «físico», entonces se preferirá el término señalado (agradezco la observación de Gustavo Ortiz-Millán).

como parte de este grupo aquellas teorías que se denominan: materialismo (reductivo, eliminativo, revisionista), monismo anómalo y teoría de la identidad (de tipo y de caso).

Materialismo y fisicismo

El materialismo es la teoría de que todo en el mundo es de carácter material. Se opone al dualismo pues afirma que o no existe algo así como el alma o, en todo caso, ésta no es otra cosa más que el resultado de la organización de la materia. El materialista, por lo tanto, rechaza parte del problema planteado por Descartes al no ver la necesidad de buscar una relación entre dos sustancias, pues, en última instancia, sólo existe la sustancia material. Pero, por otro lado, sigue estando dentro del esquema de pensamiento cartesiano al continuar hablando de *sustancias*.

Podemos encontrar antecedentes materialistas en épocas tan remotas como en las antiguas mitologías de las grandes civilizaciones o en los filósofos físicos de la época presocrática. No obstante, el antecedente más cercano lo tenemos en la misma época de Descartes, en lo que se llamó *materialismo mecanicista* de los siglos XVII y XVIII: Pierre Gassendi, Thomas Hobbes, John Locke, Francis Bacon, De La Mettrie.⁷ Este materialismo floreció sobre la confianza producida por los éxitos de las predicciones físicas a partir de las leyes de Newton. Se concibió entonces al universo como configura-

⁷ La lista por supuesto, puede alargarse todo lo que se quiera dependiendo de la importancia que la tesis materialista tenga dentro del autor a considerar. Algunos candidatos más serían, por ejemplo, Giordano Bruno o Galileo Galilei.

do por un determinismo causal fuerte del que nada podría escapar, incluido el libre albedrío humano.⁸

Mención aparte merecen los autores considerados dentro de la corriente del materialismo dialéctico o histórico. Pensadores como Marx, Engels o Lenin son materialistas por el énfasis que ponen en las condiciones materiales de producción para la comprensión de la dinámica social más que por una postura ontológica respecto del mundo. Sin embargo, la extensión de la tesis materialista dialéctica de estos autores a menudo se tradujo en teorías materialistas de corte epistémico y ontológico en otras áreas como en la filosofía de la mente.⁹

En filosofía de la mente se conocen como materialistas a un conjunto de teorías recientes sobre el problema mente-cuerpo: materialismo reduccionista, materialismo eliminativista y monismo anómalo. No obstante, la comprensión de estas teorías requiere de la explicación previa de algunos conceptos que a continuación veremos.

Finalmente, por materialismo podemos entender la presunta postura que la ciencia en general tiene sobre el mundo en la actualidad. Sin embargo, en rigor, esta postura de la ciencia quedaría mejor representada por el término «fisicismo».

⁸ Postura bien representada por la afirmación determinista de Laplace (1820): «Una inteligencia que en un instante dado supiera todas las fuerzas que actúan en la naturaleza y la posición de cada objeto en el universo —si estuviese dotada de un cerebro suficientemente vasto para hacer todos los cálculos necesarios— podría describir con una sola fórmula los movimientos de los mayores cuerpos astronómicos y los de los átomos más pequeños. Para tal inteligencia, nada sería incierto, el futuro, como el pasado, serían un libro abierto».

⁹ Hay que recordar que la tesis con la que Marx obtuvo su doctorado en Jena es una comparación entre los sistemas físicos de Demócrito y Epicuro.

En una primera distinción, el fisicismo se diferencia del materialismo al afirmar que el mundo está compuesto por, y sólo por, las entidades que postula la ciencia física contemporánea: tanto materia, como energía. Esta distinción parece superflua, pero no lo es. Mientras que el materialismo en su expresión más básica sólo hace una afirmación ontológica sobre el tipo de cosas que pueblan el mundo, el fisicismo hace una afirmación también disciplinaria y epistémica sobre la forma en que podemos acceder a ese mundo y a su explicación. El fisicista se compromete con un tipo de explicación que resultaría necesaria y suficiente sobre todo fenómeno que podamos experimentar: materia, energía y leyes que las gobiernan.

El término es introducido originalmente por Carnap y Neurath, filósofos del Círculo de Viena, durante la década de los años 30 para referir a la tesis de que el lenguaje del conocimiento debería de ser, en su estrato más general y básico, de carácter físico. No se trataba, en su origen, de una tesis ontológica sino sobre el lenguaje de la ciencia. Los fisicismos posteriores se desarrollaron en otras direcciones que desbordaron la mera interpretación lingüística de la tesis original. Como se puede observar, el fisicismo, al igual que el materialismo en su forma más general, no es propiamente una tesis sobre la relación mente-cuerpo, sino sobre el mundo en general; es la tesis metafísica de que todo en el mundo es físico o, en otras palabras, que no existe propiedad en el mundo que no pueda ser explicada en términos físicos. De aquí se puede derivar la afirmación de que todo en el mundo es *explicable* (en principio) por la ciencia física (actual o posible).

De este modo tenemos entonces, una teoría que intenta explicar las propiedades del mundo, cualesquiera que estas sean, las mentales incluidas, a partir de las propiedades que

describe la física. En el fisicismo, esta relación entre propiedades toma una forma lógica específica conocida como *superveniencia*. Existen múltiples formulaciones distintas de esta relación; una forma de definirla es:

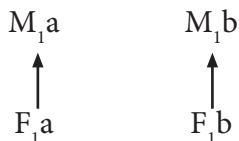
Decir que una propiedad X *superviene* a una propiedad Y significa que dos objetos que tengan la misma propiedad Y , tendrán la misma propiedad X ; y de igual manera: dos objetos que difieran en sus propiedades X tendrán diferentes propiedades Y .

Se puede observar que la dependencia es de X hacia Y : la presencia de las propiedades X depende de la presencia de las propiedades Y . El siguiente esquema explica ambas relaciones:

No hay diferencia en M sin
diferencia en F



No hay igualdad en F sin
igualdad en M



La superveniencia es una forma común de entender las relaciones entre niveles teóricos en la ciencia. Así, se puede afirmar que las propiedades moleculares supervienen a las atómicas, o que las propiedades celulares supervienen a las moleculares. En el ámbito que nos interesa, la relación de superveniencia afirmarí que no hay diferencia entre dos objetos en sus propiedades mentales sin que difieran en sus propiedades físicas. Así, los estados mentales son *dependientes de* los estados físicos, es decir, de los estados neurales que las

realizan.¹⁰ En forma rápida podemos resumir entonces la tesis de la superveniencia de lo mental a lo físico de la siguiente manera: *No hay diferencia mental sin diferencia neural y no hay igualdad neural sin igualdad mental.*¹¹

Fisicismo tipo y fisicismo caso

Para comenzar con este apartado, aclaremos de inicio la distinción entre tipo (*type*) y caso (*token*). El retruécano de Luis de Góngora «Cuando pitas flautas, cuando flautas pitas» contiene seis casos (*tokens*), pero sólo tres tipos (*types*). De la misma manera, el número 21122 contiene cinco casos o dígitos (*tokens*) pero sólo dos tipos (*types*).

La tesis de la superveniencia divide al fisicismo en dos: el *fisicismo tipo* y el *fisicismo caso*. Hay que notar que la superveniencia establece una dirección clara en la relación entre propiedades. La identidad entre las propiedades Y (neurales) implica la identidad entre las propiedades X (mentales), mientras que la diferencia entre las propiedades X (mentales) implica la diferencia entre las propiedades Y (neurales). Esta dirección, como vimos, implica que X se *reduce* a Y . Se trata

¹⁰ Si bien es cierto que esta *dependencia* es de carácter lógico, sólo hay un paso para la afirmación reductiva: los estados mentales *se reducen a* (no son otra cosa más que) estados físicos.

¹¹ Al ser una relación lógica, los filósofos suelen argumentar a favor o en contra de la superveniencia construyendo experimentos mentales a partir de mundos posibles. El reto está en encontrar un mundo posible en donde alguna de las relaciones supervenientes sea falsa. Experimentos mentales que son clásicos en esta área son los de los *zombies filosóficos* (seres con propiedades físicas idénticas, pero sin estados de conciencia) o el del *espectro invertido* (dos sujetos con la misma fisiología, pero con diferente experiencia cualitativa del espectro de colores).

de la relación entre dos conjuntos de propiedades: las del *tipo* x y las del *tipo* y. Dentro de estos conjuntos de propiedades podemos establecer subconjuntos. Así, dentro del conjunto de todos los estados con propiedades mentales podemos indicar estados mentales con propiedades específicas: estados de dolor, de tristeza, de percepción, de recuerdo, de imaginación, etcétera. El fisicista entonces puede afirmar que el *tipo* de los estados dolorosos es idéntico al *tipo* de estados cerebrales en donde se activan determinado grupo de conexiones neuronales (*fibras C*, por ejemplo). Esto no compromete al fisicista a afirmar qué dolor específico se asocia con cuál conjunto de activaciones neuronales en particular. Lo que sí es cierto, afirma la tesis del fisicismo tipo, es que todos los estados de y que se reducen a los estados de x (estados de dolor, por ejemplo) son de determinado tipo neural. Esto deja lugar a la posibilidad de que un dolor en específico esté implementado por una red de conexiones neurales distinta en dos personas diferentes. De igual manera, la activación de determinada región cerebral asociada con el dolor no necesariamente producirá dos dolores idénticos en dos personas distintas. Sin embargo, en todos estos casos, ni los dolores ni las zonas de activación neural en cuestión dejan de formar parte de sus tipos específicos. A esta forma de fisicismo se le conoce también como teoría de la identidad de tipo (*Type Identity Theory*).¹²

Por otro lado, tenemos al fisicismo de caso, también conocido como teoría de la identidad de casos (*Token Identity Theory*) o monismo anómalo.¹³ Esta tesis afirma que cada

¹² Véase: U.T. Place, 1956; H. Feigl, 1967 y J.C.C. Smart, 1959.

¹³ Véase: Donald Davidson, 1980.

caso de estado mental *es idéntico a* un caso de estado neural. Para el monismo anómalo no existen leyes generales que correlacionen causalmente los tipos de estados dolorosos con tipos de estados neurales (de ahí el adjetivo *anómalo*). Sin embargo, los estados dolorosos se reducen, en efecto, a sus correspondientes neuronales (de ahí al término de *monismo*). En el monismo anómalo se conserva la identificación de tipos de estados mentales (dolor, percepción, creencia, etcétera), pero su correlación neural no tiene por qué pertenecer a un mismo tipo. De tal modo, no se pueden establecer correlaciones «legales» estables entre tipos de propiedades. Debido a esta resistencia a la posibilidad de encontrar constantes legales entre propiedades mentales y físicas, se ha considerado que el monismo anómalo ni siquiera es un caso de *fisicismo*.¹⁴

Al apoyar al *fisicismo tipo* y rechazar al *fisicismo caso*, la tesis de la superveniencia abre la posibilidad a otra teoría de la reducción en la filosofía de la mente: el *funcionalismo*. Regresaremos a esta tesis y a las teorías que posibilita en el último apartado de este capítulo. Por lo pronto, falta por revisar un tipo de *fisicismo* aún más radical.

Materialismo reductivo y eliminativo

Abordemos la cuestión de la relación mente-cerebro desde otra perspectiva. Todos sabemos que tenemos una mente en forma directa e intuitiva. No se trata de un conocimiento proposicional ni adquirido por vía de otras personas. La mente se conoce en primera persona en su mismo ejercicio. Sabemos cómo se sienten los dolores y, de hecho, somos ca-

¹⁴ Véase: J. Kim, 1996.

paces de especificar cualidades muy detalladas de diversos tipos de sensaciones dolorosas. Lo mismo con las emociones, sentimientos, deseos, creencias, percepciones, etcétera. Si se nos pide, podemos dar múltiples características de cada uno de estos estados mentales, pues los conocemos por haberlos experimentado. Podemos describir nuestros dolores como intensos, insoportables, «sordos» o agudos; nuestros deseos como apremiantes o pasajeros; los recuerdos como efímeros, borrosos o nítidos; los sueños como inquietantes, «muy reales» o «significativos»; y así sucesivamente. Sin embargo, esas características serán siempre de tipo «interno», de tipo cualitativo o, como gustan los filósofos en llamar: de tipo *fenoménico*. No hay nada en mi experiencia interna que me indique la existencia de neuronas, conexiones sinápticas o neurotransmisores. Para todo fin práctico, una persona bien puede pasar por la vida con una rica experiencia mental sin tener idea de que posee un cerebro. No obstante, ese es exactamente el lenguaje que usa la ciencia para describir nuestras experiencias mentales: neuronas, sinapsis, módulos, neurotransmisores, potenciales de acción, etcétera. Si el fisicismo es cierto, entonces mi experiencia fenoménica y la descripción científica de ella deberían de ser compatibles. La pregunta entonces es, ¿cómo es posible conciliar dos descripciones de lo que presuntamente es lo mismo?

Existen dos respuestas a esta pregunta: la *reductiva* y la *eliminativa*. Para el materialismo reduccionista¹⁵ el lenguaje usado para describir nuestra experiencia fenoménica puede ser entendido como una forma de psicología que, a diferencia de la psicología científica, sería de carácter popular. Esta

¹⁵ Véase: Place, 1956 y Smart, 1959.

psicología popular emplea términos vagos y no bien definidos que se han consolidado en el lenguaje cotidiano por medio de la costumbre. Su uso tiende a hacernos caer en falacias (reificaciones o hipóstasis) o en incomprensiones sobre la estructura y funcionamiento de nuestra mente. Pero tales errores se resolverían (y muchos de los problemas asociados se disolverían) si comprendiéramos que la mente no es más que cerebro, por lo que los conceptos del lenguaje popular con el que nos referimos a la mente pueden ser reducidos por completo y sin pérdida de significado al lenguaje de la ciencia.

Quizá no nos gusta pensar en nuestras emociones y deseos en términos de neuronas y sinapsis; la idea causa rechazo por ser contraintuitiva, pero la historia de la ciencia es la historia de la lucha contra nuestras intuiciones más arraigadas: a nadie le resulta intuitivo (hasta la fecha) que la tierra se mueva, que el calor sea el resultado del promedio del movimiento de las partículas de la materia o que el tiempo esté en relación con la velocidad de los cuerpos que se desplazan en el espacio. Y sin embargo, así es. Los reduccionistas afirman que, de la misma manera que aprendimos a identificar el agua con su composición química y el calor con el movimiento promedio de las moléculas, así nos terminaremos por percatar que los estados mentales *no son otra cosa más que* activaciones de grupos neuronales. Así como la termodinámica fenoménica fue reducida por la termodinámica estadística, de la misma manera la psicología popular deberá ser reducida (por completo y sin pérdida de significado) por la neurociencia.

Pero esa reducción, en el mejor de los casos, es un deseo: conseguir una neurociencia completa y terminada capaz de comprender hasta el último reducto del funcionamiento ce-

rebral. Mientras eso no suceda (y es dudable que así sea), la reducción que desea el materialista entre ambos lenguajes deberá esperar.¹⁶

El matrimonio Churchland (Patricia y Paul)¹⁷ sabe de las dificultades asociadas a la tesis de la reducción, por lo que han propuesto otra forma de materialismo quizá más radical: el eliminativo. No se trata, dicen ellos, de reducir los conceptos de la psicología popular al lenguaje de las neurociencias, sino eliminar por completo el uso de dichos conceptos. La psicología popular es una descripción fallida que en ningún sentido ha colaborado a una mejor comprensión de qué es la mente. Por tal motivo, no hay necesidad alguna de intentar correlacionar en forma reductiva los conceptos psicológicos con los científicos. La correspondencia tiene sentido cuando se logra hacer ver que el concepto *x* usado por una teoría no era otra cosa más que el concepto *y* de otra, pero no cuando una teoría se muestra como falsa a favor de otra. A nadie se le ocurriría, por ejemplo, decir que el concepto de brujería fue reducido por el de psicosis o que el flogisto fue reducido por la teoría química contemporánea. Aquellos conceptos simplemente cayeron en desuso por mostrarse como descripciones incorrectas de la realidad.

¹⁶ El materialista reductivo se enfrenta además a otras objeciones como la que emana de la violación de la ley de Leibniz: dos cosas que son idénticas tienen todas sus propiedades en común $\{(\forall x) (\forall y) [(x \leftrightarrow y) \leftrightarrow (\forall F) (Fx \leftrightarrow Fy)]\}$. Es claro que los estados mentales presentan propiedades en primera persona que no podemos adjudicar a las neuronas y sus sinapsis; de igual manera, por más que yo crea que mi mente no es más que cerebro, no tiene sentido decir que mi mente pesa 1.3 kg.

¹⁷ Véase: Patricia Churchland, 1989 y Paul Churchland, 1999.

No es que los fenómenos que dieron origen a la idea de brujería, del flogisto o del calórico hayan desaparecido, sino que los hemos aprendido a reinterpretar a partir de otro marco conceptual de referencia. Eso es lo que hace la ciencia. y eso es justamente, dicen los Churchland, lo que deberemos hacer en el caso de la ciencia de la mente. Por supuesto, esto implica una exigencia similar a la que implicaba el materialismo reductivo: deberemos tener una ciencia completa del sistema nervioso central con la finalidad de evidenciar como inútiles los conceptos psicológicos de modo que podamos reinterpretar toda nuestra vida mental a partir de neuronas y sinapsis. Lo que estos autores proponen en realidad, más que una teoría es una especie de programa revisionista: ahí en donde se vayan logrando avances en nuestra comprensión del mecanismo de la mente, debemos aceptar la eliminación de los conceptos psicológicos correspondientes y aprender a ver la mente desde el nuevo esquema propuesto por la ciencia. Es posible que, en el camino, algunos de tales conceptos que usamos cotidianamente se vean validados... aunque parece poco probable.

Conductismo lógico: la mente es conducta

Cada quien conoce sus propios estados mentales por la evidencia directa. Mi conocimiento de mi estado de enojo, de preocupación, de asombro o de duda es idéntico a su existencia. En el instante mismo en que sucede mi percepción existe una evidencia directa de su ocurrencia para el propio sujeto. Lo mismo parece cumplirse para el resto de los estados mentales. Por supuesto, siempre es posible encontrar ocurrencias mentales que pasan desapercibidas para la aten-

ción de la persona. En eso justamente se basan terapias como el psicoanálisis: en hacer evidente lo que desconozco de mi propia vida mental, sin embargo, incluso ese proceso de traer a la conciencia atenta los estados mentales *subconscientes* parece llevarse a cabo en forma introspectiva, es decir, interna.

Ahora bien, ¿cómo es que conozco la existencia de los estados mentales ajenos? En este caso la experiencia (la ocurrencia del estado) no coincide con su conocimiento. Lo que conozco de otro sujeto no es su estado de enojo, sino la conducta que lo evidencia. Gracias a la conducta (verbal o corporal) puedo asumir, con mayor o menor grado de precisión, que la otra persona se encuentra en un estado de dolor, explicar en retrospectiva lo que dicha persona estaba pensando cuando llevó a cabo determinada acción o predecir lo que ese otro sujeto pensará (y de qué forma actuará) cuando reciba cierta información. De lo único de lo que tengo evidencia respecto de la vida mental de los otros es de su conducta.¹⁸

La afirmación básica del conductismo establece: lo único empíricamente constatable de la vida mental es la conducta. Como antes hicimos con la tesis de la superveniencia, la re-

¹⁸ Este es el problema filosófico de las otras mentes: dado que no tengo acceso directo a las mentes ajenas, y dado que la conducta por sí misma no justifica en forma deductiva la existencia de una mente, ¿cómo sé que los otros tienen estados mentales como los que yo experimento en mí mismo? Esta pregunta abre el debate en torno al solipsismo que se encuentra presente como un paso en la cadena argumentativa de la duda metodológica cartesiana y que tiene importantes consecuencias escépticas. Hoy en día, el problema de las otras mentes es abordado por múltiples disciplinas más allá de la filosofía: psicología cognitiva, psicología social, neurociencia cognitiva, antropología. El reciente descubrimiento de las neuronas espejo ha traído nuevas perspectivas al debate. Al respecto, véase el segundo capítulo de este libro.

lación lógica implícita en el conductismo podría formularse así: *no hay diferencia entre dos estados mentales a menos que pueda señalarse una diferencia conductual.*

El conductismo, en realidad, más que una única teoría coherente puede describirse como un conjunto de teorías que comparten una actitud similar hacia la forma de comprender y estudiar la mente. La obviedad de la asimetría que mencionamos entre nuestra forma de acceder a nuestros propios estados mentales y la forma en que conocemos sobre los estados mentales ajenos hace que podamos hablar de una cierta actitud conductista cotidiana. Ahí, por supuesto, no existe teoría alguna, solo una asunción implícita que resulta eficaz en nuestra vida cotidiana para comprender y explicar las acciones de los otros. Como programa de investigación, en cambio, el conductismo cobró importancia a partir de la segunda década del siglo xx y fue de primordial influencia hasta mediados de dicho siglo, cuando surge con fuerza la ciencia cognitiva.¹⁹ No obstante, hoy en día sigue siendo un programa de investigación vivo y fecundo en la psicología, tanto de corte experimental como terapéutico.

Es necesario distinguir dos tipos de conductismo: el psicológico o metodológico y el conocido como filosófico, analítico o lingüístico. En su primera versión, el conductismo es

¹⁹ Podemos reconocer el inicio del declive conductista en el *Review* (1959) que Chomsky hizo sobre el libro de Skinner *Verbal Behavior* (1957). En dicho libro, Skinner intentaba explicar la adquisición del lenguaje en los niños únicamente a partir de los postulados conductistas; Chomsky demuestra por medio de argumentos como el de la *pobreza de estímulo* que dichos postulados no son suficientes para explicar la adquisición por parte del niño de una habilidad tan compleja como la lingüística en un periodo tan breve.

una tesis metodológica sobre la ciencia de la psicología: la psicología es una ciencia empírica que trata con indicadores observables, por lo que su objeto de estudio es la conducta, no la mente. John Watson, uno de los más destacados conductistas, afirmaba: «La psicología, como la ve el conductista, es una rama experimental pura, estrictamente objetiva, de la ciencia natural. Su objetivo teórico es la predicción y el control.» (1913, p. 158). Por lo tanto, todas las fuentes de obtención de información del psicólogo son externas, observables y cuantificables. Así, la ciencia de la conducta no tiene por qué asumir ninguna hipótesis respecto del carácter ontológico o funcional de la mente, sino sólo debe preocuparse por establecer correlaciones empíricas entre estímulos y respuestas a partir de modificadores tales como refuerzos positivos o negativos.

El conductismo psicológico, en el sentido aquí definido, toma frecuentemente la forma de un programa de control y modificación de la conducta ya sea con fines terapéuticos o pedagógicos. Se trata de condicionar las respuestas conductuales por medio de los estímulos adecuados. A partir de esta idea, encontramos los dos programas de condicionamiento más importantes: el clásico, cuyo principal representante es Ivan Pavlov, y el operante que tiene en B. F. Skinner a su principal exponente. En este condicionamiento se trata de motivar la conducta a partir de la asociación entre diversos estímulos. El ejemplo típico es el experimento pavloviano de desplazamiento de la conducta de salivación de los perros desde la imagen de la comida hacia el sonido de la campana. En el condicionamiento operante, en cambio, se motiva la conducta y se modela mediante reforzadores positivos y negativos; es decir, se modela la respuesta asociándola a las

consecuencias de la acción del propio sujeto. Ejemplo de este tipo de condicionamiento es la conducta que exhibe una rata al pulsar una palanca que da comida y evitar pulsar aquella que le ha dado un choque eléctrico en el pasado.

Por otro lado, tenemos al segundo tipo de conductismo: el lógico o analítico. El conductismo lógico afirma que *lo que significan* los términos mentales que utilizamos *no es más que* conducta o disposiciones conductuales. En ese sentido, el conductismo lógico, al igual que el psicológico, no se compromete con una tesis de carácter ontológico, sino que hace una afirmación sobre la semántica de nuestro lenguaje mental. En última instancia, para ambos tipos de conductismo, el problema mente-cerebro es un pseudoproblema que se diluye una vez que reconocemos que no es a una entidad inmaterial (la mente) a lo que nos referimos cuando usamos nuestro vocabulario mental, sino sólo a la conducta observable o posible (queda abierto el problema de decidir si tal entidad realmente existe). En realidad, el conductismo lógico no es una tesis sobre la mente, sino sobre nuestro lenguaje mental. Entonces, para este tipo de conductismo todo término mental puede ser reducido sin pérdida de significado a conducta real o posible (disposiciones conductuales):

$$\{(\forall x) [Mx \leftrightarrow (\alpha x \bullet \beta x)].$$

Donde:

M es un estado mental determinado

α es el conjunto de expresiones conductuales (corporales o verbales) exhibidas.

β es el conjunto de disposiciones conductuales para el estado mental en cuestión.

Esta expresión bicondicional muestra el carácter reductor de la tesis conductista. No obstante, falta por aclarar aun el significado de la expresión «disposición conductual».²⁰ Un predicado es disposicional cuando es posible que se presente ante un conjunto de condiciones determinadas. La cualidad descrita por un predicado de este tipo no está «en acto», sino sólo como potencialidad. Lo que un predicado disposicional señala, en realidad, es un enunciado de carácter condicional: si se presentan las condiciones x , entonces se presentará la cualidad y . Un ejemplo de predicado de este tipo es el de *fragilidad*: que el vidrio de mi ventana sea frágil significa que bajo ciertas circunstancias (un golpe de determinada fuerza), el vidrio se romperá. El vidrio no pierde su cualidad de frágil por no haberse roto nunca. Otro tanto puede decirse de predicados como soluble, resistente, inflamable, impermeable, etcétera.

En el ámbito que nos interesa, el de la conducta, un predicado es disposicional si refiere a la conducta posible dado un contexto determinado. Predicados conductuales de carácter disposicional serían: valiente, cobarde, iracundo, tenaz, gracioso, etcétera. Lo que todos estos términos señalan son enunciados condicionales que relacionan una determinada circunstancia con un conjunto de conductas posibles. Por ejemplo, afirmar que «Claudia es una persona tenaz en su trabajo» significaría que ante circunstancias adversas en su medio laboral (escasez de recursos, un jefe

²⁰ Véase: G. Ryle, 2005.

autoritario, compañeros desleales, etcétera), exhibirá conductas como: no renunciar, no abandonar la encomienda asignada, encontrar vías de solución para los problemas que se presentan, etcétera.

El conductismo lógico tiene sus raíces en el programa del positivismo lógico. De tal programa, toma dos ideas principales: la reducción interteórica, y el principio verificacionista del significado.²¹ La idea positivista de la reducción consiste en la posibilidad de traducir un lenguaje teórico en otro más básico sin pérdida de significado. Claramente, en el caso de la psicología, el lenguaje a reducir es el relativo a los términos mentales (deseos, creencias, emociones, etcétera), mientras que el lenguaje reductor es el relacionado a la conducta observada (real) y observable (posible). Tenemos entonces, la misma idea reductora presente también en las teorías materialistas revisadas previamente.

Al mismo tiempo, los positivistas ligaban el significado de las proposiciones a la posibilidad de su verificación: una proposición es significativa (tiene sentido) si y solo si es posible establecer las condiciones empíricas de su verificación. En el caso de la ciencia de la mente, los únicos indicadores empíricos disponibles (los que hacen propiamente que tal ciencia pueda ser empírica) son los conductuales. Por tanto, las proposiciones sobre la mente tienen sentido únicamente

²¹ Aparte del citado libro de Ryle, podemos referir a la lectura de Carnap «Psychology in a Physical Language» (1932), Hempel «The Logical Analysis of Psychology» (1949) y las Investigaciones Filosóficas de Wittgenstein (1953). Sobre el programa en general del positivismo lógico se puede consultar la compilación clásica de Ayer (1965) y el reciente y amplio estudio de Stadler (2013).

si se demuestran como expresiones referidas al comportamiento cuantificable y manipulable del sujeto.

El conductismo resultó ser un programa teórico de gran impacto debido principalmente a las siguientes razones: 1) la consideración del problema tradicional mente-cuerpo como un pseudoproblema, dejando así, sin efecto la cuestión sobre la ontología de la mente, 2) la presencia de cierta intuición obvia a nivel cotidiano sobre la imposibilidad de conocer en forma directa la mente ajena, 3) la ventaja que planteaba la identificación de elementos cuantificables para la consolidación de la psicología como una ciencia empírica y 4) la viabilidad de la modificación efectiva de la conducta a través del descubrimiento de regularidades causales entre estímulos, respuestas y reforzadores.

No obstante, la fuerza de su influencia metodológica en la psicología solo es comparable con su declive teórico como explicación satisfactoria de los estados mentales. Esto se debe a que el conductismo lógico, en tanto teoría que intenta encontrar la cualidad necesaria y suficiente detrás de nuestro vocabulario mental, se encuentra con algunos problemas que parecen insuperables. Mencionaré solo dos de ellos: 1) la lista de conductas reales y posibles que definirían cada estado mental resulta inagotable, de modo que no resulta verosímil pensar en un conjunto finito y exclusivo de estados conductuales que reduzcan al término mental en cuestión y que sean propios sólo de ese término y 2) la definición de tales términos mentales nunca se libra de otros términos mentales, pues siempre que se intenta especificar la conducta propia de un estado mental aparecen otros términos que también requerirían en sí mismos de una reducción conductual... ¡Y así sucesivamente! Estos problemas ponen

en entredicho el programa del conductismo lógico y abren paso a la última teoría reductiva que presentaremos en este apartado: el funcionalismo.

Funcionalismo: la mente es una función

En su intento por llevar a cabo un estudio científico de la mente, el conductismo dio la espalda por completo a lo que parece más propio de la vida mental: la inmediatez de la experiencia subjetiva. Lo que sucedía al interior de la mente (si acaso sucedía algo) era visto por el conductista metodológico como una caja negra. En correspondencia, para el conductista lógico los términos mentales que no permiten su reducción por medio de términos conductuales carecen de significado: son términos vacíos o pseudoconceptos. La mente, como tal, no jugaba dentro de la teoría conductista ningún rol causal y, si así lo hacía, tal rol no resultaba relevante para el estudio empírico.

En principio, por tanto, la hipótesis conductista resulta viable para cualquier autómeta carente de vida mental pero cuya conducta responda a las regularidades descritas por el conductismo para la definición de cada estado mental correspondiente. Pensemos en el caso de los actores profesionales: son capaces de simular con precisión dolores, emociones o creencias por medio de su conducta sin encontrarse en realidad en el estado mental correspondiente. O a la inversa, pensemos por ejemplo en el caso de una comunidad de *súper-espartanos* educados en la resistencia al dolor y en la supresión de todo gesto que evidencie que se encuentran

sufriendo.²² No se trata de que el súper-espartano en cuestión no sienta dolor, sino que suprima la conducta que lo delate. En ese caso, tendríamos estado mental sin conducta evidente. Si ambos casos son cuando menos posibles, entonces se demostraría que la identidad entre estados mentales por un lado y conducta real y posible por otro sería falsa. El problema proviene, justamente, del desprecio del conductista por lo que sucede dentro de la caja negra de la mente. Una teoría que intente evitar estos problemas debería sugerir una forma de «entrar» a dicha caja negra para definir la dinámica mental y tomarla en cuenta en las descripciones causales de la conducta del sujeto.

El funcionalismo original, propuesto por Hilary Putnam a finales de la década de 1960,²³ surge como producto de estas críticas a la tesis conductista. A esta tesis le llamaremos *funcionalismo computacional*. La afirmación es que los estados mentales *no son otra cosa más* que estados funcionales. ¿Qué son los estados funcionales? Para responder a esta pregunta es necesario introducir el concepto de *máquina de Turing* en el cual se basa Putnam para proponer el concepto de *autómata probabilístico*. Una máquina de Turing es un dispositivo idealizado creado por Alan Turing en la década de 1940 para resolver un problema matemático propuesto originalmente por David Hilbert.²⁴ En el dispositivo de Tu-

²² El ejemplo es de Putnam en *Cerebro y conducta* (1983).

²³ Además del artículo referido en la nota anterior, el texto donde Putnam propone su funcionalismo es «La vida mental de algunas máquinas» (1981).

²⁴ Véase: A. Turing, «On Computable Numbers, With an Application to the Entscheidungsproblem» (2004); R. Penrose, *La mente nueva del emperador* (2002); J. Gray, *El reto de Hilbert* (2005).

ring se pueden introducir datos (*input*) para obtener ciertos resultados (*output*) de acuerdo con las instrucciones de la máquina y a la situación interna (estados internos) en que se encuentra. Este dispositivo es totalmente determinista en el sentido de que dado cierto *input* introducido al sistema y el estado interno en que éste se encuentre, sabremos siempre con absoluta certeza cuál será el *output* obtenido y el estado interno subsiguiente. Además, se trata de un dispositivo que afirma relaciones causales: los inputs causan que el dispositivo se encuentre en determinado estado, el cual, a su vez, es causa de los outputs que el dispositivo ha de emitir dadas las relaciones en que dicho estado se encuentre respecto de otros estados del sistema. La transición entre estados internos viene dada por un algoritmo que técnicamente se llama *tabla de máquina*. Así, una máquina de Turing puede tener una tabla de máquina que le indique: si recibes el input I_1 y te encuentras en el estado interno E_1 , entonces emites el resultado O_1 y pasas al estado interno E_2 .

Putnam (1971) toma la idea de una máquina de Turing y la modifica para proponer un modelo bajo el cual cabría pensar cómo funciona la mente humana. A este modelo le llama un *autómata probabilístico*. Cabe suponer que la mente humana funciona bajo principios probabilísticos. Putnam entiende por esto dos cosas distintas: 1) la mente puede evaluar probabilidades de estados de cosas y actuar maximizando la utilidad de acuerdo con la teoría racional de la decisión y 2) las respuestas conductuales y los estados mentales de una persona, aunque los podamos inferir o predecir con cierta certeza, se encuentran indeterminados. Así pues, un modelo computacional que se ajuste a la mente humana será una máquina de Turing modificada: un autómata probabilístico.

Este autómata estará provisto de órganos sensoriales para el ingreso de información (input), órganos motrices (output) y estados internos cuya transición esté especificada por un algoritmo probabilístico del tipo: si recibes el input I_1 y te encuentras en el estado interno E_1 , entonces *prefiere* (con un determinado grado de probabilidad) la respuesta O_1 y *prefiere* (con un determinado grado de probabilidad) pasar al estado E_2 . Los estados internos E son entonces estados que se definen de acuerdo con su relación probabilística causal con los datos de entrada, los datos de salida y con otros estados. Es decir, para cada estado mental E_n hay un conjunto, y sólo uno, $\{I_n, O_n, R_n\}$ que lo define, en donde I son los datos de entrada, O los datos de salida y R las relaciones causales que mantiene con otros estados mentales. E_n queda entonces definido como un estado funcional del autómata probabilístico.

A partir de esta caracterización del funcionalismo es fácil ver que un autómata como el descrito puede ser instanciado en un organismo biológico (como el ser humano), en algún otro sistema físico (una computadora) o, incluso, en un sistema no físico. En caso de que la analogía con la mente humana sea correcta, esto equivale a decir que los estados mentales *son* estados funcionales. Un dolor, como ejemplo de estado mental, queda definido funcionalmente como un estado del sistema *normalmente* ocasionado por un daño ocurrido en su estructura física (input), que da origen a otros estados internos (creencia de que se está sufriendo un daño, deseo de retirar la parte afectada, etcétera) y a la *inclinación* espontánea a retirar la parte afectada (output) de la fuente que causa el dolor.

Hay que aclarar algunos puntos importantes sobre el funcionalismo computacional:

1. El funcionalismo computacional, a diferencia del conductismo lógico, admite la existencia de estados internos del sistema y de sus poderes causales. El hecho de que el sistema exhiba determinada conducta es consecuencia de la existencia de cierto estado interno que lo produce; de tal forma es solucionable el problema de la posible existencia de los súper-espartanos: casos en que el sujeto no exhiba una conducta normalmente asociada con determinado estado mental y, no obstante, estemos dispuestos a adjudicar la posesión de tal estado por parte del sujeto en cuestión. En esto consiste parte de la crítica del funcionalismo computacional contra el conductismo lógico: los estados mentales no son disposiciones conductuales, en realidad, son causa de tales disposiciones.
2. Aunque el funcionalismo es compatible con el fisicismo, en el sentido de que se da por sentado que los dispositivos físicos que instancian autómatas probabilísticos están sometidos a las leyes físicas, sin embargo, el funcionalismo computacional de Putnam *no es* un fisicismo en el mismo sentido del fisicismo-tipo. Y no lo es debido a la tesis de la *múltiple realización* de los estados funcio-

nales.²⁵ Lo que importa para caracterizar los estados mentales son las relaciones que éstos guardan con otros estados mentales, así como con los datos de entrada y de salida; es decir, lo que importa es su rol funcional y no la base material en que se realizan. De hecho, el funcionalismo se presenta como una refutación al fisicismo-tipo en dos sentidos: a) dar una descripción completa de un estado mental es dar su estado funcional y esto no dice nada sobre su estado físico y b) para que el fisicismo-tipo sea cierto, es necesario que se dé la equivalencia: «Hay un estado mental *X* si y sólo si hay estimulación de las fibras *c*»; sin embargo, el funcionalismo afirma que un estado mental (el dolor, por ejemplo) considerado como un estado funcional puede ser realizado en múltiples dispositivos físicos diferentes que no necesariamente han de corresponderse con las fibras *c*. Dar una definición funcional de un estado mental implica especificar *todas* sus relaciones funcionalmente pertinentes, pero entre estas relaciones causales no figurarán aquellas que el estado en cuestión mantenga con el dispositivo físico en que se realiza. De esta manera, el funcionalismo supone que la instancia

²⁵ La tesis de la *múltiple realización* que es central a este tipo de funcionalismo afirma que es posible la instanciación del mismo estado funcional en diferentes organizaciones físicas. Si identificamos estado mental con estado funcional eso significa que es posible, en principio, encontrar sistemas con cualidades mentales en sustratos físicos distintos de los organismos que somos los seres humanos, por ejemplo: en otros animales, computadoras, toda clase de robots, seres extraterrestres o, incluso, seres incorpóreos (toda vez que el estado funcional no exige ninguna relación causal con el sustrato físico).

física no es un elemento necesario para incluir dentro de la definición de la mente.

3. Putnam señala explícitamente que la afirmación de la analogía (o eventual identidad) existente entre la mente humana y un autómata probabilístico no pasa de ser una *hipótesis verosímil* (mucho más, quizá, que las hipótesis materialista y conductista) que ha de ser comprobada o refutada empíricamente, pero que no se trata en absoluto de un argumento deductivo contundente. Y esto es así por dos razones: a) aunque parezca bastante verosímil que la mente humana funciona bajo leyes probabilísticas, también parece cierto que en las situaciones cotidianas el ser humano no actúa siempre de acuerdo con la teoría de la acción racional, además, b) el modelo esbozado de un autómata probabilístico parece ser en extremo simplificado y racional como para dar cuenta de la enorme complejidad de la mente humana. Sin embargo, dice Putnam, esto último no es por sí mismo un defecto, pues el modelo podría ser complejizado lo suficiente como para semejar lo más posible a la mente humana, pero, por lo pronto, basta con que dicha mente se comporte en términos generales como lo indica el modelo para que las conclusiones extraídas sean verosímiles. Será necesario esperar a la elaboración de una teoría psicológica desarrollada que sea capaz de mostrar que los estados mentales se pueden definir por completo por medio de su rol funcional.

Es necesario mencionar que en la actualidad existen múltiples formas de funcionalismo que se han desarrollado a partir de

la propuesta original de Putnam. Estos funcionalismos se diferencian por su grado de aceptación de la tesis de la múltiple realización, de modo que algunos de ellos se asimilan con tesis fisicistas.²⁶ Debido a la plasticidad de la tesis funcionalista y a su asociación originaria con el procesamiento computacional, el funcionalismo ha logrado sobrevivir incluso como forma estándar de comprensión de la mente. En la conciencia popular, tanto como en la investigación empírica (en neurociencia, psicología cognitiva o en ingeniería computacional) se toma como verdadera la analogía computacional de la mente. Esta analogía entre mente y computadora está detrás de la teoría computacional representacional de la mente que revisaremos a continuación.

2. La perspectiva clásica de la mente

Los tres programas teóricos revisados hasta ahora intentaban dar respuesta al problema mente-cerebro replanteado en términos de la definición de los conceptos mentales a partir de su reducción a algún otro conjunto de estados: el estado físico, el estado conductual o el estado funcional. Aunque se trata de una actualización del problema cartesiano original de la comunicación entre la sustancia mental y la física, la atención ya no giraba en torno al aspecto ontológico del problema. Esto se demuestra con la compatibilidad ontológica del conductismo con posturas dualistas o monistas por igual o con la tesis explícita de la múltiple realización que está en la base del funcionalismo.

²⁶ Véase: S. Shoemaker, «Some Varieties of Functionalism» (1981) y a García Carpintero, «El funcionalismo» (2007).

Las tres posturas teóricas en cuestión resultaron fecundas de diversas maneras. El fisicismo se ha convertido en una tesis estándar implícita en el trabajo de la mayoría de los científicos, tanto en las ciencias físicas como en las neurociencias. El conductismo, como hemos revisado, tuvo profunda influencia en la psicología empírica y terapéutica en donde sigue siendo un programa fértil de investigación. Por su parte, el funcionalismo introdujo la analogía computacional con lo que colaboró a la consolidación de la perspectiva más influyente en las investigaciones del cerebro y de la mente a partir de la década de 1960: la *Teoría Computacional Representacional de la Mente* (TCRM).

La TCRM es el esquema teórico detrás de la interdisciplina que surgió hacia mediados del siglo xx y que fue conocida como ciencia cognitiva. Este nuevo campo de investigación se encontraba comprometido con el estudio empírico de la mente, en particular, de los procesos relacionados con la cognición: toma de decisiones, razonamiento, lenguaje, memoria, aprendizaje, percepción. En esta tarea se integraron disciplinas como la lingüística formal, el funcionalismo filosófico, la psicología cognitiva, las ciencias de la computación, la teoría de la información y, posteriormente, disciplinas como la neurociencia cognitiva,²⁷ la economía y la teoría de la decisión.

Como suele suceder, la TCRM es un nombre que sirve para agrupar un número amplio de posturas tanto filosóficas como científicas al interior de las cuales no deja de

²⁷ La neurociencia cognitiva surge en la década de 1980 con el *Journal of Cognitive Neuroscience* y con la edición del texto fundacional *Wet Mind. The New Cognitive Neuroscience* de Kosslyn y Koenig.

haber fuertes debates e inconsistencias. Además, se trata de un campo muy dinámico que ha visto modificados sus postulados teóricos conforme se avanza en el conocimiento al interior de cada disciplina. Hay que decir también que el término de *ciencias cognitivas*²⁸ expresa un interés conjunto de investigación más que una interdisciplina homogénea y unificada en metodologías, vocabularios y perspectivas teóricas. Tenemos por ejemplo que, tradicionalmente, se integra dentro de tal empresa a la Inteligencia Artificial, la cual, en realidad no es una disciplina teórica como el resto, sino un proyecto tecnológico con un objetivo específico: la replicación (o imitación) de la conducta inteligente en un dispositivo no orgánico.

Finalmente, hay que señalar una distinción importante: tenemos por un lado la TCRM, que es un conjunto de postulados teóricos, y por el otro, tenemos a las ciencias cognitivas, que es un campo de trabajo interdisciplinario. El surgimiento de ambas está entrelazado históricamente: la analogía entre mente y computadora sirvió tanto para alentar el desarrollo de las ciencias de la computación como para el estudio empírico y filosófico de la relación entre cerebro y mente. Actualmente, a la luz de los logros tecnológicos en materia de computación, parece obvia la asociación entre inteligencia, mente y computadora. No obstante, la TCRM es un programa teórico que puede ser desafiado, puesto a prueba o, eventualmente, substituido sin que el área de investigación

²⁸ El término en sí mismo es objeto de debate: si se utiliza en singular se apunta hacia una ciencia unificada con métodos, técnicas e hipótesis teóricas bien definidas; en cambio, en plural, señala hacia el ámbito de trabajo en común de diversas disciplinas que conservan su independencia teórica y metodológica.

de las ciencias cognitivas quede sin vigencia. Esto lo han intentado hacer los teóricos de la corporización cuyas ideas revisaremos en el siguiente capítulo. Pero, incluso, frente a estas nuevas ideas, quienes proponen una perspectiva computacional-representacional de la mente permanecen convencidos de sus postulados. Esta convicción está claramente expresada por Paul Thagard cuando afirma:

En términos teóricos, la idea más fructífera ha sido el enfoque computacional-representacional de la mente [...] La hipótesis central de las ciencias cognitivas es que la mejor forma de estudiar el pensamiento es entenderlo en términos de estructuras de representaciones mentales sobre las que operan procesos computacionales. [...] Sin duda, el modelo es el mejor a nivel teórico y experimental entre todos los enfoques del estudio de la mente que se han desarrollado hasta el momento. (Thagard, [2005], 2010: 28)

A pesar de la variedad teórica que integra la TCRM, podemos identificar una serie de postulados que suscribe de alguna forma toda teoría perteneciente a esta perspectiva; no todos ellos se encuentran en las teorías del espectro computacional-representacional, pero cuando menos algunos han de ser afirmados para identificar a tal propuesta como parte de la TCRM:

1. Existen estados mentales internos y su estudio es necesario para la comprensión del funcionamiento de la mente (a diferencia de lo que afirma el conductismo).
2. El conocimiento es procesamiento de la información. Las funciones de la mente son principalmente cognitivas; la mente, por lo tanto, es un dispositivo que procesa información.

3. La mente crea representaciones a partir de los datos que obtiene del medio.²⁹
4. Las representaciones mentales son manipuladas (computadas) por medio de reglas sintácticas innatas o adquiridas (ya sean deductivas o probabilísticas) que forman algoritmos.
5. El resultado de las manipulaciones simbólicas puede ser expresado por medio de conducta.
6. El modelo abstracto de procesamiento de la información puede ser instanciado de igual modo por cualquier dispositivo (biológico o artificial) que respete las relaciones funcionales señaladas por el modelo (tesis funcionalista de la Múltiple Realización).
7. El cuerpo (o el *hardware*) es el medio de obtención de la información relevante del ambiente y es el vehículo de expresión y ejecución de los resultados de las computaciones mentales.

Detrás de estos siete postulados hay dos afirmaciones básicas. Como el nombre mismo lo indica, la TCRM integra dos tesis generales sobre el funcionamiento y la estructura de la mente: 1) la mente funciona a partir de manipulación sintáctica de símbolos sobre los que se opera por medio de reglas lógicas y 2) dichos símbolos son representaciones del medio en que se desenvuelve el organismo. Ambas tesis son estrictamente independientes, lo que significa que es posible sostener una sin necesidad de suscribir la otra. Una teoría que integre ambas será una teoría fuertemente cognitivista,

²⁹ Uno de los principales problemas de la TCRM (y de las ciencias cognitivas en general) es la gran variación en el significado del término «representación». Este tema será abordado hacia el final de este capítulo.

lo que significa que estará menos dispuesta a aceptar afirmaciones rivales como las corporizadas. Sin embargo, la aceptación de una u otra de esas tesis abre un abanico teórico muy amplio que a continuación explicaremos en forma general.

La tesis computacional

Para los computacionalistas, la mente es una *Máquina Universal de Turing* de carácter probabilista. Esto quiere decir que es un dispositivo que puede manipular símbolos mediante diferentes subrutinas (reglas en la tabla de máquina) para lograr distintos propósitos por medio de reglas de operación no deterministas (probabilistas o, como dicen los matemáticos: *estocásticas*).

Como hemos mencionado, esta tesis surge por primera vez en el trabajo original de Alan Turing (1936), para después ser retomada como programa filosófico por Hilary Putnam (1967). Posteriormente, autores como Jerry Fodor (1975), Allan Newell y Herbert Simon (1976) o Zenon Pylyshyn (1984) colaboran a la consolidación de la idea de que la mente lleva a cabo computación. En un principio, dicha computación se entendió como la manipulación simbólica a partir de reglas sintácticas de carácter deductivo. Esto significa que, si las condiciones iniciales están definidas y las reglas son claras y completas, entonces siempre es posible llegar a un resultado determinado, siempre y cuando dichas reglas sean aplicadas correctamente. Por esta vía se crearon programas expertos (que tenían un objetivo único y específico) que corrían algoritmos cuyos resultados eran siempre lógicamente necesarios. Newell y Simon (1956), por ejemplo, desarrollaron algoritmos deductivos por los cuales fueron

capaces de demostrar 38 de los 52 teoremas que contiene el *Principia Mathematica* de Russell y Whitehead.

En sus inicios, toda el área de investigación de las ciencias cognitivas entendía que los procesos mentales más importantes, aquellos que capturaban la esencia de lo mental, eran los de carácter cognitivo. Es por esto que, los algoritmos computacionales se desarrollaron exclusivamente para modelar funciones como el razonamiento deductivo, la toma de decisiones, el almacenamiento y recuperación de un ítem dentro de un listado (aprendizaje y memoria), etcétera. El desarrollo de estos programas expertos fue lo que caracterizó a la primera forma de la Inteligencia Artificial, aquella que fue bautizada por John Haugeland como GOFAI: *Good Old Fashioned Artificial Intelligence* (Inteligencia Artificial a la Vieja Usanza). El objetivo de esta IA era replicar la inteligencia humana en un mecanismo. Esto demostraría que, en efecto, la inteligencia *es* computación. Para ello el Test de Turing (1950)³⁰ fue la herramienta principal de contrastación de la tesis. Quizá el logro más importante de esta forma de IA se obtuvo en 1997, cuando los ingenieros de IBM desarrollaron la super-computadora *Deep Blue*, la cual fue

³⁰ El test propuesto por Turing funciona de la siguiente manera: un sujeto recibe mensajes (por un monitor, por ejemplo) de dos interlocutores. Uno de ellos es una persona y el otro es una computadora. El sujeto tiene que identificar quién es cada interlocutor. Si el sujeto se equivoca en tal asignación la misma cantidad de casos en que acierta, entonces estará adivinando a quién pertenece cada mensaje, con lo que se puede decir que la computadora presenta conducta inteligente.

capaz de vencer al campeón de ajedrez Gary Kaspárov.³¹ Hasta el día de hoy, con todo y las críticas de múltiples teóricos sobre la pertinencia del Test de Turing para evaluar la presunta inteligencia de una computadora, existen torneos internacionales para premiar al software que sea capaz de pasar alguna versión del test.

La tesis computacional, como se ha descrito hasta aquí, sólo requiere de manipulación sintáctica por medio de reglas. Cualquier sistema que pueda hacer esto será considerado un sistema con propiedades mentales (o de modelación mental). De aquí se infiere que a esta tesis no le son relevantes dos cosas: el sustrato físico en que se implementa el algoritmo y el contenido de los símbolos manipulados. Hay que recordar que el funcionalismo computacional tiene como principal tesis la de la múltiple realización.

Actualmente, sin embargo, se han diversificado las formas de computacionalismo. Esto se debe al avance en todos los campos de conocimiento involucrados: desde las ciencias de la computación hasta las neurociencias. Y es precisamente el campo de la *neurociencia computacional* en donde se restringe la tesis de la múltiple realización. Mucho del trabajo que hoy en día se hace en neurociencia cognitiva pasa por la elaboración de modelos computacionales basados en el

³¹ El resultado final fue de 3-2 a favor de Deep Blue con un empate. Previamente, en 1996, una versión anterior del programa fue derrotada por Kaspárov 4-2. Kaspárov, hay que decirlo, jamás estuvo de acuerdo en que su derrota se debiera únicamente al poder computacional del programa, pues siempre sospechó la manipulación humana de algunas movidas por parte de su contrincante.

funcionamiento de redes neuronales.³² Dado que lo que se busca es la comprensión del funcionamiento de un sistema biológico determinado, y no de cualquier sistema con un rol funcional similar, entonces dichos modelos deben utilizar variables que tomen en cuenta aspectos relevantes de la constitución biológica de la mente humana.³³ Así, entre más comprometidos estamos con la descripción de nuestro propio sistema biológico como fundamento de la mente, menos fuerza adjudicaremos a la distinción entre rol funcional y sustrato físico. Esto, por sí mismo, ya representa uno de los desafíos importantes que enfrenta el computacionalismo, el cual es conocido como *problema de la implementación*.

Por otro lado, tenemos el problema del contenido de los símbolos manipulados. Para el computacionalismo radical la sintaxis es necesaria y suficiente para definir la función mental. No obstante, esto acarrea algunos problemas. El más conocido de ellos es el que ha denunciado John Searle (1980) en su famoso *argumento del cuarto chino*. El argumento está basado en el siguiente experimento mental: imaginemos que no sé una palabra de chino. Por alguna razón, alguien me encierra dentro de un cuarto que cuenta con dos ventanas (una de ingreso y otra de salida). Frente a mí tengo una mesa de trabajo con un conjunto de garabatos. Además, cuento con un listado que relaciona algunos de esos garabatos con otros. Se me dan las siguientes instrucciones: «cuando recibas un papel con una cadena de garabatos, consulta tu listado y arma la cadena correspondiente con los garabatos sobre la mesa,

³² Una referencia importante sobre el uso de herramientas y conceptos computacionales aplicados al estudio de procesos y redes neuronales es el de Patricia Churchland y Terrence Sejnowski (2017).

³³ Véase Patricia Churchland, 1986; Tappenberg, 2010.

finalmente, entrega el resultado por la ventanilla de salida». Imaginemos que dichos garabatos, tanto los que recibo como los que entrego, son de hecho símbolos chinos con significado (para un hablante chino). Para mí, esos símbolos no son tales pues, de hecho, no me significan nada. Ni siquiera los reconozco como representaciones de algo. Sin embargo, para una persona ubicada afuera del cuarto, parecería que quien está dentro es un hablante competente del idioma chino. La virtud lingüística (semántica), sin embargo, no está en mí (pues no sé una palabra de chino), sino en el listado que estoy siguiendo al pie de la letra y que, en todo caso, estará en un idioma que sí entiendo. Lo único que yo estoy haciendo realmente es manipulación sintáctica de algo que, para mí, no son signos. Está claro que para llevar a cabo la tarea no requiero acceder al contenido (la semántica) del garabato en cuestión. El argumento de Searle se cierra diciendo: la computadora hace exactamente eso. Yo, como ser que posee una mente, sí accedo al contenido de los símbolos que utilizo, luego entonces, ni la computadora entiende nada (por lo que es absurdo el objetivo de la IA) ni la mente realiza computaciones cuando funciona.

El argumento de Searle es un clásico dentro de la filosofía de la mente y ha sido objeto de múltiples debates y críticas. No obstante, señala un punto de enorme importancia: parece que una teoría de la mente no puede dejar de lado las propiedades semánticas que aparentan ser tan esenciales a la función mental. Si queremos entender la forma en que los símbolos manipulados tienen significado entonces nos tenemos que comprometer con una teoría semántica de la mente. Y eso implica abordar el tema de la representación.

La tesis de la representación

El experimento mental de Searle pone en cuestión la suficiencia de la tesis computacional y señala hacia la necesidad de que una teoría de la mente integre las cualidades semánticas de ésta. Esto quiere decir que no basta la organización de los símbolos que son manipulados por el sistema, sino que hace falta ver cada símbolo como una representación con un contenido que lo individualiza. Quienes afirman que la idea de representación es esencial en una teoría de la mente, piensan que es justamente ese contenido el que «liga» a la mente con el mundo. Si no fuera así, sería posible que el sistema mental llevara a cabo algoritmos deductivamente coherentes e infalibles en términos computacionales, pero que no tuvieran relación alguna con el mundo en el que el sujeto se desenvuelve, haciéndolo por completo ineficaz en su acción cotidiana.³⁴

Asumimos entonces que los estados mentales son representaciones de «cosas». Pero de inmediato vemos que esta tesis presenta varias cuestiones filosóficas a resolver: ¿qué tipo de «cosas» exactamente son posible contenido de una representación?, ¿cómo se «liga» la representación con su contenido?, ¿cómo es posible conciliar la tesis de la representación mental con una perspectiva científica del mundo sin necesidad de aludir a cualidades o sustancias de carácter no-físico? Y, en primer lugar, ¿qué es exactamente lo que estamos entendiendo por *representación mental*?

³⁴ Este tema se relaciona con el debate epistemológico entre el *fundacionismo* y el *coherentismo* como forma de justificación de las creencias. Al respecto, véase la compilación de García, C. L., A. Eraña, y P. King (2012).

En efecto, el término de «representación» se nos ofrece en forma inmediata y muy natural al hablar del funcionamiento de la mente, pero no por eso es poco problemático. El uso habitual del término nos condiciona a pensar que tenemos claro su sentido, pero nada más lejos de la realidad. Tradicionalmente, la filosofía ha entendido a la representación en forma preminentemente lingüística-conceptual. Los conceptos son formas de representar objetos o clases de objetos: manzanas, casas, instituciones, relaciones de parentesco, cualidades físicas, especies biológicas, etcétera. De hecho, es posible establecer una taxonomía de dichas clases con base en su generalidad hasta llegar a las clases de objetos cuyo contenido está especificado en la forma más general posible. Tal es el caso de las tablas de las categorías desde Aristóteles hasta Kant. Pero también se habla de representación en la percepción: la imagen que observo es representación de la escena, el sonido de la alerta sísmica representa un sismo inminente, etcétera. Otro tanto ocurre con las reglas y prohibiciones: tal símbolo representa la prohibición de estacionarse; con los mapas como representaciones esquemáticas del territorio; con las fotografías o pinturas con respecto a la escena que presentan; con los planes como representación de un curso de acción a seguir, etcétera.

Entonces, el problema es que cuando hablamos de representaciones mentales podemos referirnos a: pensamientos, conceptos, percepciones, ideas, impresiones, reglas, esquemas, imaginaciones, recuerdos, creencias, deseos, palabras, mapas... o, incluso, neuronas, módulos cerebrales, potenciales de acción o sincronía en los disparos

sinápticos.³⁵ Resulta difícil encontrar un significado claro y unívoco para un concepto que debe poder integrar tal variedad de aplicaciones.

En la segunda mitad del siglo XIX, el filósofo aristotélico Franz Brentano se da a la tarea de buscar la cualidad propia de la mente que la diferencia de cualquier objeto material (Brentano, 1926). Brentano encuentra que tal diferencia radica en lo que llamó *intencionalidad*: la cualidad de la mente de *dirigirse hacia* algo ajeno a ella misma.³⁶ Para Brentano, esta característica era necesaria y suficiente, es decir: no existe ningún estado mental que no sea intencional (todos se dirigen hacia algo que es su contenido), y no hay nada intencional en el mundo que no sea de carácter mental (los estados físicos no son intencionales). Un asunto que se presta a un intenso debate es el de saber si ambas partes del bicondicional que enlaza la mente con la intencionalidad son verdaderas. Hay quien afirma que hay estados mentales (como las sensaciones puras o la ansiedad) que no tienen un contenido intencional; a la inversa, hay quienes piensan que es posible encontrar formas de intencionalidad no mental en la naturaleza: por ejemplo, en el movimiento del girasol hacia la fuente de luz.

³⁵ Sobre la variedad de significados de «representación», véase: Thagard (2010), Johnson-Laird (2004), McDermott (2001) y Pinker (2005).

³⁶ Husserl, discípulo de Brentano, retoma la tesis de la intencionalidad y la desarrolla hasta fundar a partir de ella su *filosofía fenomenológica*. Esta tradición filosófica de descripción de los estados mentales será dominante en la filosofía continental durante el siglo XX, pero se mantendrá ajena al desarrollo de la filosofía de la mente de corte analítico-científico que aquí estamos describiendo.

Es importante mencionar que, bajo la idea brentiana, el contenido puede ser cualquier cosa o estado de cosas en el mundo hacia el cual refiera un estado mental. De hecho, el contenido puede ser un estado mental en sí mismo: como cuando recordamos un sueño o una percepción. Así, el contenido puede ser el libro que percibo en estos momentos o la paz mundial, la cual es contenido de mi deseo. Por su parte, el estado mental en sí mismo es la forma en que el sujeto se relaciona con tal contenido, sin mediación de ninguna otra entidad. Es decir, no hay necesidad de postular «representaciones intermedias» al estilo de los datos de los sentidos³⁷ que sirvan para explicar la forma en que sujeto y mundo se relacionan. Esto, como puede verse, tiene el beneficio de evitar explicar la ontología y el funcionamiento de dichos intermediarios, pero deja abierta la explicación sobre la forma del «dirigirse hacia». Tal dirección puede ser descrita en la forma en que se presenta (por medio de herramientas fenomenológicas, por ejemplo), pero para una filosofía que pretende integrar la comprensión científica del vínculo, es necesario brindar una explicación causal de la relación entre

³⁷ Los datos de los sentidos o «*sense data*» es la entidad intermedia entre la mente y el mundo postulada por filósofos británicos como Russell, Ayer y Moore (aunque pueden encontrarse antecedentes en Locke y Hume). Esta entidad tenía la cualidad de la cercanía con la mente, lo cual los hacía incorregibles, pero mantenían una distancia con el mundo que los hacía una representación falible, lo que permitía explicar casos como percepciones falsas o las ilusiones. Esta teoría ha sido objeto de múltiples críticas, por ejemplo: se ha objetado que impide el contacto directo con la realidad física; se piensa que fomenta una teoría poco parsimoniosa de la mente, postulando entidades no demostrables o se objeta que para poder cumplir su función, deberían tener las mismas cualidades que los objetos físicos que representan, lo cual sería absurdo.

mente y mundo. El abanico de las teorías semánticas que abordan este tema es también bastante amplio.

Ahora bien, si la primera parte de la tesis de Brentano es cierta (todo estado mental es intencional), entonces tendríamos dos partes de la relación: el estado mental y su contenido intencional. Bajo esta distinción podemos hacer clasificaciones de los estados mentales de acuerdo con su forma de dirigirse hacia su respectivo contenido: hay una forma específica en que el deseo se dirige hacia el objeto deseado. Lo mismo puede decirse de la creencia, la percepción, el recuerdo, etcétera. De la misma manera, podemos rastrear la constancia de un mismo contenido intencional a lo largo de diversos estados mentales: aunque no sea lo mismo desear, pensar, aborrecer o soñar una hamburguesa, el contenido en cada uno de esos casos es siempre el mismo. Una descripción similar es la que se proponía Brentano en su psicología empírica, y es la tarea que retoma Husserl en forma de descripción de esencias de la conciencia por medio de su psicología fenomenológica. No obstante, como ya hemos mencionado, está por verse que en efecto la intencionalidad sea la marca esencial de todo estado mental.

Por otro lado, la segunda parte de la tesis de Brentano (que solo los estados mentales son intencionales) abre nuevamente el debate sobre la viabilidad de la IA. Si admitimos la posibilidad de que un mecanismo no orgánico tenga mentalidad, entonces, por la tesis de Brentano, tendrá capacidades semánticas y no solo sintácticas. Pero si Searle tiene razón en las conclusiones que extrae del experimento mental del cuarto chino, entonces es imposible que un mecanismo meramente sintáctico acceda a

ninguna forma de contenido, con lo que, según la tesis de Brentano, dicho mecanismo carecería de mentalidad.

Pero incluso afirmando que los estados mentales tienen un contenido intencional, cabe preguntar sobre las características de dicho contenido. Hay dos posibilidades: el contenido es conceptual o cualitativo. Que el contenido sea conceptual quiere decir que puede ser expresado en términos proposicionales. Hay estados mentales que claramente tienen la característica de tener una proposición asociada a ellos; tal es el caso, por ejemplo, de las creencias: mi creencia en que la Revolución Francesa inició en 1789 tiene una proposición como su contenido («la Revolución Francesa inició en 1789»). Lo mismo sucede con estados como el deseo, el pensamiento o el recuerdo. Sin embargo, hay otros estados que parecen tener una característica que escapa a la descripción proposicional: la experiencia de ver el mar por primera vez, el sonido de una trompeta o el olor del café por la mañana tienen una característica cualitativa no reducible a conceptos. Por más versado en descripciones mentales que esté un niño, difícilmente logrará hacer comprender a su hermano menor, que no se pudo subir al juego mecánico, *cómo se siente* tener dicha experiencia. Menos aún logrará replicar dicha sensación en su hermano con la simple descripción. Ese «cómo se siente» ha sido nombrado, desde Nagel (1974) como el *carácter subjetivo*

de la experiencia o, para usar una etiqueta breve de origen latino, *qualia*.³⁸

El dolor, por ejemplo, ha sido el estado mental favorito de los filósofos para abordar el tema de los *qualia*: la mayor parte de su contenido parece ser esa cualidad negativa de desagrado que nos hace al dolor tan aborrecible. De hecho, estamos tan familiarizados con tal cualidad que, con base en sus distinciones, podemos identificar y diferenciar con mucha fineza diversos tipos de dolores. De lo que no somos tan capaces es de traducir dichas diferencias cualitativas en palabras: para ello recurrimos a analogías o comparaciones más o menos burdas. Si es cierto que el contenido cualitativo no es reducible a conceptos, entonces requerimos de una teoría que integre los aspectos semánticos de los estados mentales, tanto como los cualitativos.³⁹

Las tesis de la computación y de la representación forman en conjunto el marco teórico principal que ha seguido la ciencia cognitiva desde sus inicios para brindar explicaciones de los procesos mentales. Para finalizar, presentaremos un ejemplo clásico de explicación cognitivista de la percepción visual de Marr y Poggio (1976). Este ejemplo nos servirá en

³⁸ Hoy en día, hay filósofos (Kriegel, 2006; Levine, 2006) que distinguen *qualia* de *carácter subjetivo de la experiencia*. Mientras que el primero indica la cualidad de la experiencia: aquella gracias a la cual soy capaz de diferenciar entre el color rojo y el sonido de una trompeta, el segundo es la característica de un estado mental de pertenecer a un sujeto. Se dirá más al respecto en el tercer capítulo.

³⁹ Sobre la cualidad subjetiva de la experiencia, se puede consultar la compilación de Ezcurdia y Hansberg: *La naturaleza de la experiencia. Volumen 1: Sensaciones* (2003).

el siguiente capítulo para comparar las explicaciones de la TCRM y la teoría de la corporización.

Lo que los autores intentan explicar específicamente es la *disparidad estereoscópica*. Se le conoce así al proceso por el cual el cerebro forma una sola imagen, de la que somos conscientes, a partir de las dos imágenes ligeramente distintas provenientes de cada ojo. Marr y Poggio parten de la analogía computacional. Para ellos, las diferencias estructurales entre la mente y la computadora no son relevantes para la explicación de los procesos; en ambos casos, afirman, se puede hacer el análisis en términos de computación por medio de algoritmos.

Según esta explicación, el cerebro mide las diferencias entre las dos imágenes y utiliza esas mediciones para calcular la profundidad de la escena. El problema es que la simple información aportada por las imágenes no basta para descartar los diferentes arreglos posibles en que pueden estar ordenados los objetos en la escena presentada. Es decir, diferentes posiciones de un mismo objeto podrían dar como resultado la misma escena visual. La tesis es que, para descartar estas posibilidades y lograr discriminar el arreglo verdadero, el cerebro necesita realizar inferencias lógicas basadas en algoritmos. De tal modo, aunque la información de la escena sea insuficiente, el cerebro tampoco necesita de más pistas; basta con la aplicación por medio de un algoritmo de sus propias reglas computacionales sobre los datos obtenidos. Así, Marr y Poggio afirman haber descubierto que se trata de dos reglas: *R1: Unicidad* y *R2: Continuidad* y que éstas son obtenidas por el cerebro a partir de un par de características de los objetos físicos: 1) los objetos tienen una sola posición en el espacio y el tiempo, y 2) la materia se estructura en ob-

jetos discretos. Una vez realizado el cálculo de las diferencias entre las imágenes dispares, y aplicado el algoritmo con las dos reglas en cuestión, el cerebro es capaz de saber cuál es el arreglo verdadero de los objetos en la escena visual. Como se puede observar, esta explicación requiere entender al cerebro como un dispositivo computacional altamente eficaz, que funciona en el trasfondo de nuestra conciencia para producir nuestra experiencia visual a cada instante. Y esto, aunque resulta posible, parece poco verosímil según los críticos de la teoría computacional de la mente.

Los temas de la manipulación sintáctica, del contenido mental y de la forma en que éste se representa en la mente y en el cerebro presentan múltiples aspectos abiertos a debate. Algunos de ellos han sido mencionados aquí. Las posturas filosóficas que se encuentran dentro del espectro de la TCRM ponen el énfasis en la cognición y en el contenido proposicional en diferentes grados. No obstante, parece que no todos los estados mentales pueden ser igualmente comprendidos en estos términos. Ante estados directamente relacionados con la experiencia o con la interacción con el medio físico en donde la cognición no juega un papel principal, parece haber otras opciones teóricas más viables. En el siguiente capítulo abordaremos el espectro de las teorías corporizadas, las cuales pretenden resolver de mejor manera algunos de los problemas aquí presentados.

II. TRASCENDER LA MENTE

Tanto la filosofía de la mente durante el siglo xx como las ciencias cognitivas que surgieron sobre el esquema teórico de la TCRM son herederas del planteamiento cartesiano del problema mente-cuerpo. Descartes nos dividió en dos sustancias que luego hubo necesidad de reunir. Los monismos, reduccionismos y las teorías de la identidad son intentos por explicar cómo es que dos «cosas» son en realidad una sola. Los malabarismos teóricos para justificar tal división pueden ser complejos, pero, sobre todo, innecesarios. La TCRM tiene importantes méritos. Nadie puede negar que, en gran medida, fueron los principios basados en la analogía informática los que redirigieron la investigación psicológica al interior de la «caja negra», los que congregaron a diferentes disciplinas en torno a un mismo campo de trabajo (vago y no muy definido en sus inicios) y los que dotaron a los investigadores de una analogía útil y comprensible que, al tratar a los estados mentales como procesamiento de información, tenía la ventaja de unificar el lenguaje. Además, la concepción computacional de la mente ayudó a renovar el interés filosófico en el tema de los estados mentales y de la conciencia. Por su parte, gracias

a la modelación de sistemas expertos basados en el funcionamiento de las redes neuronales tenemos hoy más información sobre procesos como la comprensión del lenguaje, la memoria, la toma de decisiones o la solución de problemas.

Sin embargo, hay que reconocer que tal perspectiva se enfrenta a problemas que, aunque pueden ser computacionalmente abordados, parecen resolverse mejor desde posturas teóricas no cognitivistas. Además, las tesis de la computación y la representación se han convertido en una especie de *a priori* teórico para todo aquel que trabaja en el área. Incluso, esta analogía da forma a la preconcepción popular de la mente y la computadora. Esto obstaculiza que se tome una distancia crítica adecuada para valorar las virtudes y defectos de la teoría y para proponer su mejoría o eventual substitución en favor de otras tesis que se encuentren mejor posicionadas para explicar los fenómenos mentales que nos interesan.

Desde la década de 1970 surgió un conjunto de propuestas agrupadas en forma genérica bajo el nombre de *cognición corporizada* (*embodied cognition*). Estas propuestas tenían en común un desafío a las tesis básicas de la TCRM. Tal desafío no era novedoso, pues el cognitivismo de corte computacional ya había sido enfrentado por proyectos como el conexionismo o por ideas en torno a la mente que provenían de tradiciones teóricas ajenas a la filosofía analítica. Sin embargo, en el primer caso, los conexionistas continuaban procediendo por medio de la modelación matemática-computacional para el estudio de los procesos cognitivos, con lo que el proyecto no implicaba un cambio radical de perspectiva. En cuanto a las ideas provenientes de otras tradiciones, en realidad hay que decir que más que un desafío se trataba de caminos que transcurrían en forma paralela la mayor parte

del tiempo. La filosofía de la mente de corte analítico y las ciencias cognitivas en general nunca vieron en la filosofía francesa o alemana un recurso a explorar. Particularmente, la fenomenología en cualquiera de sus vertientes (Husserl, Scheler, Heidegger, Merleau-Ponty o Sartre) siempre resultó críptica y lejana para filósofos que se veían más cercanos al discurso de las ciencias que al de una filosofía que ellos consideraban metafísica. La fenomenología tendría que esperar a intentos como el de Hubert Dreyfus (1972) o el de Francisco Varela (1991) para ser considerada un interlocutor serio dentro de la discusión.

En 1974, Thomas Nagel revoluciona la discusión sobre el problema mente-cuerpo al introducir a la conciencia como elemento indispensable dentro de cualquier teoría de la mente. Como se mencionó en el capítulo anterior, para Nagel, la conciencia era sinónimo de cualidad subjetiva de la experiencia (*qualia*). Al ser un elemento cualitativo y subjetivo, los *qualia* desafiaban frontalmente las explicaciones reductivas y representacionistas.

De igual forma, un concepto de cognición cada vez más amplio proveniente de las investigaciones empíricas en etología y psicología del desarrollo, así como las críticas que siempre estuvieron presentes hacia la analogía computacional que está detrás del proyecto de la IA, colaboraron al desarrollo de teorías ajenas al cognitivismo tradicional. Gracias a estos desafíos el propio cognitivismo se diversificó, surgiendo teorías que sustentan sólo un mínimo de tesis funcionalistas, representacionistas o computacionales (Dennett, 1991, 2005). Si en un principio, las disciplinas fuertes que guiaban la investigación sobre los procesos cognitivos eran la lógica, las matemáticas y las ciencias de la computación y la información,

ahora se metieron a la discusión las ciencias biológicas, la psicología del desarrollo y las ciencias de la patología mental.

De inicio, hay que señalar que el nombre de *cognición corporizada* difícilmente señala una teoría homogénea; ni siquiera refiere a un conjunto claro y unívoco de tesis sobre la cognición. Como se verá a lo largo de este capítulo, al interior de este grupo encontramos posturas distribuidas en un espectro teórico muy amplio. Más que una teoría, podríamos decir que se trata de un campo teórico de propuestas con principios afines y un objetivo común, pero sin leyes, vocabulario, ni métodos unificados; una especie de *programa de investigación* animado por intuiciones similares (Shapiro, 2007). Clark y Chalmers (1998) afirman que el rasgo común es un *anti-aislacionismo post-cartesiano*, es decir, siempre subyace la afirmación de que la conciencia no es una entidad subsistente aislada del resto del mundo (el cuerpo incluido) que produce por sí misma toda la actividad mental. Bajo este abanico encontramos por igual teorías que afirman que la cognición, y la mentalidad en general, es una propiedad sensorio-motriz que involucra al cuerpo completo del sujeto (cognición corporizada: Varela *et al*, 1991); también hay quienes afirman que los tipos de estados mentales son tales por la interacción del sujeto con el medio y esta interacción es siempre de carácter contextual (cognición situada: Brooks, 1990); finalmente, hay posturas que afirman que el contenido de los estados mentales no está determinado por sus condiciones intrínsecas (lo propiamente «mental»), sino que involucra también las herramientas de las que nos valemos

culturalmente (teoría de la mente extendida: Clark y Chalmers, 1998; Bartra, 2007).¹

De la misma forma que hicimos con el cognitivismo, podemos hacer ahora el puntuado de ideas que, sin ser necesarias o suficientes, animan la vía de investigación corporizada. Se entiende que en una teoría cualquiera dentro del espectro corporizado no tienen por qué estar presentes todas las ideas listadas a continuación, sin embargo, en principio ninguna teoría de este tipo podría pertenecer al programa si no afirma cuando menos alguna de tales tesis.²

1. La cognición siempre es una actividad que se lleva a cabo en contextos particulares y que varía en función de ellos.
2. El cuerpo y sus interacciones con el medio cumplen un papel constitutivo en los procesos cognitivos.
3. La idea de «proceso» describe mejor que la de «estado» las facultades cognitivas de un organismo. Tales procesos cognitivos, por tanto, siempre son de carácter dinámico y están enfocados hacia fines prácticos en donde el organismo interactúa con los elementos en su entorno.

¹ Una revisión introductoria sobre los conceptos de mente extendida, enactiva, situada y otros similares que forman parte del mismo programa se puede encontrar en Rowlands, 2010.

² Mientras que autores como Shapiro (2011) y Clark (1998) resaltan las diferencias teóricas entre la cognición corporizada y la cognición situada, Gallagher (2008) afirma que, aunque son distintos programas, son obvias sus conexiones esenciales al grado de que es imposible que una teoría de la cognición situada no sea también una teoría de la corporización.

4. El organismo no es pasivo en la obtención de información perceptiva, sino que la obtiene activamente del medio por vía de procesos sensorio-motrices sin necesidad de mediación simbólica ni de computaciones.
5. Los procesos cognitivos co-evolucionaron con el cuerpo para adaptarse a contextos específicos de modo que, estudiando la historia evolutiva del organismo, así como su nicho ecológico, podemos explicar tales procesos.
6. El énfasis en la importancia del estudio del organismo en su medio para entender los estados mentales representa un nuevo enfoque frente al cognitivismo tradicional, lo cual implica el desarrollo de nuevas técnicas, metodologías, vocabulario y perspectivas teóricas para la investigación.

El programa de la corporización, frente al cognitivismo de la TCRM, pretende llevar a cabo un desplazamiento teórico del centro a la periferia: es decir, del centralismo de la cerebralización computacional a la periferia de la corporalidad del organismo y su medio. Además, implica una ampliación de los límites del objeto de estudio: ya no se presenta a la cognición como la única o esencial función de la mente, sino que se expande la investigación para abarcar los procesos volitivos o emocionales, así como sus relaciones dinámicas. A esto es a lo que alude el título de este capítulo: presentaremos un proyecto que intenta *trascender* la idea tradicional de mente en un sentido intensivo (al ampliar las facultades mentales que se estudian) tanto como en un sentido extensivo (al considerar al organismo y su medio). Para ello, primero hablaremos de las ideas originales que sirven de antecedente

a las teorías contemporáneas en el área. El resto del capítulo servirá para exponer algunas de las principales tesis del proyecto corporizado: la teoría de la enacción, la cognición extendida, la mente situada o la reflexión sobre el hallazgo de las neuronas espejo.

1. Organismo y medio: la psicología ecológica

La corporización contemporánea en la filosofía de la mente tiene su antecedente más directo en la psicología ecológica. Son dos libros de J.J. Gibson los que establecen una forma alterna al computacionalismo para comprender la cognición: *The Senses Considered as Perceptual Systems* (1966) y *The Ecological Approach to Visual Perception* (1979). El interés de Gibson es proponer una explicación de la percepción visual más parsimoniosa (con menos elementos teóricos) y aplicable a las condiciones reales de los organismos que la que propone el cognitivismo. Pero para entrar a esa comparación, primero es necesario sentar las bases teóricas de las que parte la psicología ecológica.

La TCRM propone una forma de dualismo: la que separa al organismo con su medio. Para la TCRM, el cuerpo y el medio en que se desplaza el sujeto no son más que el medio contingente de adquisición de información y de ejecución de la acción. Tal contingencia está claramente expresada en la tesis de la múltiple realización. La percepción, por tanto, es algo que sucede por completo «internamente» por medio de un sistema computacional que realiza cálculos, hipótesis e inferencias. La acción, y el cuerpo que la realiza, son elementos contingentes del proceso cognitivo. Esta separación entre organismo-medio y entre percepción-acción es la que

desafía la psicología ecológica de Gibson. Para él, la unidad mínima de análisis es la conformada por el sistema organismo-medio. Deshacer este sistema en sus partes implica dejar de comprender la percepción, la cual es uno de los principales aspectos de la cognición. Es por eso que, la TCRM se ve obligada a especular respuestas que, aunque lógicamente posibles, resultan prácticamente inviables. En efecto, la visión al igual que todas las demás formas de percepción, es un proceso (y no un estado) dinámico que surge por la interrelación entre los sistemas sensoriales y motrices del organismo y su medio.

De la misma manera que los sistemas físicos, el sistema organismo-medio presenta cualidades emergentes, es decir, propiedades que solo son concebibles en el funcionamiento del sistema como un todo. Las propiedades del sistema organismo-medio son las *disponibilidades (affordances)*. Por este concepto se han de entender las propiedades del mundo referidas a las posibles acciones de un organismo dada su constitución corporal. Así, una acción X es disponible (se encuentra en el *pool* de posibilidades motrices) para un organismo Y en un ambiente Z. Según el concepto de disponibilidad, toda la información que el organismo requiere para su acción ya está presente en el medio. No es necesario realizar inferencias especulativas a partir de trozos de información que han de ser conectados por medio de reglas sintácticas. La información, en el sentido de la psicología ecológica, no es «procesada», sino recuperada del acoplamiento estructural entre organismo y medio.

Ahora bien, no toda la información está disponible para todos los tipos de cuerpos. La información disponible y el *pool* de acciones posibles dependen del acoplamiento entre

el organismo y su medio. Y esto, a su vez, es un producto de su historia evolutiva. Los organismos tienen una historia evolutiva en medios específicos que los ha vuelto aptos para obtener la información necesaria para su sobrevivencia. Dos organismos distintos en el mismo medio responden ante distintas disponibilidades. Gibson afirma que estas características en el medio provocan respuestas inmediatas sin mediación de representación alguna. Vemos directamente en la silla un objeto para sentarnos, pues nuestra propia estructura corporal lo permite, no así para un perro o un gato; un ave puede ver el árbol como fuente de protección mientras que otro organismo lo ve como alimento.

La diferencia de la teoría ecológica respecto de la TCRM radica en que entre el organismo y el medio hay una relación directa. No es necesario postular ningún intermediario para hacer accesible la información al sujeto, pues esta se obtiene por completo en forma contextual. Además, esta relación directa arroja una nueva luz sobre la forma de ver la percepción y la acción: como dos aspectos del mismo proceso. La percepción sin ejecución motriz en el medio resulta en sensaciones sin significado; de igual manera, la acción solo es tal si es dirigida por la percepción, de lo contrario es solo movimiento caótico, involuntario y no dirigido. Para la psicología ecológica, la experiencia se comprende como la dialéctica percepción-acción. Que un organismo tenga determinada experiencia visual quiere decir, por lo tanto, que sea capaz de ver cursos de acción posibles que le permitan obtener mayor información por medio de la exploración.

En este punto, se puede pensar que la psicología ecológica tiene sentido cuando se trata de percepciones y acciones en el medio, pues parece resultar una explicación más acorde a

los tiempos de reacción que implican una acción eficaz en el mundo. Pero cabría suponer que cuando se trata de procesos cognitivos de mayor jerarquía de abstracción, como la toma de decisiones, la memoria, la imaginación o el razonamiento, la psicología ecológica resulta claramente insuficiente. No obstante, hay partidarios de esta propuesta (Michaels y Palatinus, 2014) que afirman que la psicología ecológica basta también para explicar ese tipo de procesos. Para ellos, *ningún* estado mental requiere de representaciones. El aprendizaje es visto como la modificación del organismo de modo tal que se encuentre mejor situado para recabar información y actuar en consecuencia gracias a las experiencias previas. No es que el sujeto cuente con un almacén central de información del que pueda disponer cuando la situación lo requiera. Una explicación de este tipo requiere postular sistemas neurales que realicen las funciones de almacenamiento y búsqueda computacional requeridos por el modelo cognitivista. En cambio, bajo la idea del acoplamiento estructural, producto de la evolución, el aprendizaje se ve como una modificación selectiva de la atención para refinar la identificación de la información que ya está en el medio. Es decir, el aprendizaje es el proceso mediante el cual el sujeto «aprende a ver» de mejor manera.

Estamos ahora sí en posición para ofrecer la explicación ecológica de la visión. Recordemos que, para Marr y Poggio,³ en la visión estereoscópica el cerebro se encuentra ante el dilema de decidir cuál es el arreglo visual que corresponda más adecuadamente con la realidad. Para ello tiene que hacer hipótesis sobre ambos modelos y decidir con base en un par de criterios. Gibson, en cambio, piensa que el organismo no

³ Véase el capítulo anterior.

tiene por qué reconstruir la realidad en un modelo,⁴ sino que navega en el medio a partir de las diferencias de luz y sombra en la escena visual. Estas diferencias acarrearán toda la información necesaria, pues la luz trae consigo información sobre la fuente que la origina y las condiciones de los objetos, por lo que no es necesario la elaboración de inferencias. Los contrastes de luz y sombra que el organismo recibe en un momento determinado son nombrados como **«arreglo óptico ambiental»**. Toda la luz de este arreglo converge en un mismo punto: las retinas. La diferencia con las explicaciones cognitivistas es que éstas conciben tal arreglo como estático. Puesto que la luz que reciben los ojos en un momento determinado puede corresponder a diferentes arreglos de los objetos en el medio, entonces se hace necesario que el cerebro realice inferencias para determinar la situación física en la que se encuentra. Pero esto es una abstracción artificial. En la realidad, los organismos se desplazan en el medio, obteniendo a cada instante información nueva. Los diferentes arreglos posibles se descartan con esta nueva información disponible. Aunque hay elementos que cambian (dimensiones de los objetos, texturas, brillos, tonalidades, ángulos, etcétera), hay otros que permanecen invariables (las proporciones de los diferentes triángulos formados por las líneas de luz que convergen en los ojos desde un mismo objeto). Son éstos últimos los que el sistema perceptivo utiliza para especificar

⁴ «El mundo es su mejor modelo» es una afirmación del ingeniero robótico del MIT Rodney Brooks (1991), que ha sido tomada como lema de la perspectiva ecológica, y que enfatiza la inutilidad de la creación de representaciones mentales (gráficas, lingüísticas o de ningún otro tipo) para la comprensión de la acción inteligente en el mundo.

las posiciones y las propiedades de los objetos en la escena. Esto sólo es posible concibiendo a un observador dinámico.

Para las teorías de la corporización, la visión se asemeja más a una película que a una fotografía, como sucedía en la perspectiva cognitivista. Los procesos perceptivos en general, son activos y extendidos, no están en el cerebro, sino que involucran al cuerpo completo, sus movimientos e interacciones con el medio. Lo dicho aquí para la visión, opera de forma similar para el resto de modalidades de la percepción, pero es particularmente aplicable a la percepción auditiva.

2. Enacción: la mente en el cuerpo

El neurobiólogo chileno Francisco Varela vio en el estudio de la conciencia un área ideal para la intersección disciplinaria. Se encontraba bien situado para tender puentes entre disciplinas distintas. Su formación profesional y su trabajo en el laboratorio al lado de Humberto Maturana le permitieron tener una sólida concepción empírica de la biología de la mente. Realizó estudios de filosofía en los que se formó en la fenomenología husserliana, interesándose particularmente por la obra de Merleau-Ponty quien aplicó la filosofía de Husserl al análisis del cuerpo vivo (*Leib*). Finalmente, vio puntos de interés para la descripción de los estados mentales en las técnicas de meditación del budismo, por lo que se convirtió en practicante de esta disciplina.⁵ El resultado de

⁵ La cercanía de Varela con el budismo lo llevó a organizar, junto con el Dalai Lama, un grupo de conferencias sobre la meditación y el estudio científico de la mente. Esta amistad con el Dalai Lama los llevó a fundar en 1991 el *Mind and Life Institute*, organización dedicada a explorar dicha relación entre ciencia y budismo.

este tránsito constante entre disciplinas, métodos y perspectivas lo materializó Varela en una propuesta de investigación de la mente al mismo tiempo filosófica y empírica: la *neurofenomenología*.⁶

El concepto central de la perspectiva teórica fundada por Varela es el de *enacción*. Hoy en día la enacción es un proyecto que se integra dentro del amplio campo del programa corporizado. Como todas las empresas al interior de dicho proyecto general, la de la enacción tiene lazos estrechos con otras ideas corporizadas. Hay un núcleo central de supuestos que son compartidos por las teorías enactivistas, extendidas y situadas. Pero los puntos de mayor interés radican precisamente en las divergencias sutiles más que en las coincidencias entre estas posturas. A su vez, el alcance de la perspectiva enactivista se ha extendido y diversificado más allá de su origen en las investigaciones de los dos biólogos chilenos, al grado de pretender convertirse en un nuevo paradigma igual de fecundo que el de la TCRM, pero sin el espíritu excluyente o reduccionista de esta última.⁷

Podemos definir la enacción como: el resultado de las interacciones dinámicas entre la fisiología del organismo, sus sistemas sensorio-motrices y el medio en el que ese organismo se desenvuelve. Se trata de poner el énfasis en el acoplamiento estructural que tienen los organismos con su medio y que, según afirma el enactivismo, es constitutivo de las cualidades cognitivas del sujeto. Ese acoplamiento estructural está sustentado en los bucles de retroalimentación (*feedback loop*) sensorio-motrices que operan incesantemente. Tales bucles son circuitos de percepción y acción que forman unidades

⁶ Véase Varela 1991 y 1996.

⁷ Di Paolo y Thompson, 2014.

funcionales. La percepción y el movimiento pueden ser diferenciados en forma analítica en la reconstrucción de sus funciones, pero en el ejercicio del organismo no son módulos autónomos ni subsidiarios. A diferencia de la TCRM, el enactivismo no da por sentada *a priori* la existencia de un sujeto y un objeto que se encuentran separados. Por el contrario, al igual que en la psicología ecológica, el organismo *enactúa* en el medio creando con ello su *mundo de experiencia vivida*. Se entiende entonces que sólo un organismo con una determinada estructura corporal puede tener las facultades mentales (procesos y estados) correspondientes. Es por ello que tal acoplamiento resulta básico para la comprensión de la cognición, pues es justo en los circuitos continuos entre la acción y la percepción en que se encuentra la característica esencial de los procesos cognitivos.⁸

Si la tesis de la enacción como acoplamiento estructural es cierta, entonces se desprende que la tesis de la múltiple realización, en donde se fundamenta buena parte de la tesis computacional, resultaría falsa. La cognición no es, como afirma el funcionalismo, el producto de un rol funcional independiente de su realización física y que, por lo tanto, puede ser instanciado en cualquier sistema orgánico o mecánico, siempre y cuando se respeten las relaciones funcionales correspondientes. Para el enactivismo, el cuerpo no solo es re-

⁸ Como se vio previamente, la psicología ecológica comparte con el enactivismo la importancia que se concede al medio y al acoplamiento estructural en los procesos cognitivos; sin embargo, para el enactivismo el elemento de mayor importancia son los bucles de retroalimentación sensorio-motrices que, como se verá más adelante en este mismo apartado, sirven de herramienta de explicación de fenómenos relacionados con la cognición a múltiples niveles: orgánicos, neurológicos, sociales.

levante, sino *constitutivo* del proceso cognitivo, es decir, que el tipo de cuerpo que se tiene determina los circuitos sensorio-motrices y, con ello, el tipo de procesos cognitivos que el sistema exhibe. Cosmelli y Thompson (2010), por ejemplo, demuestran la inviabilidad empírica del experimento mental del cerebro en una cubeta recorriendo paso a paso los requerimientos corporales que deberían cumplirse para poder tener un cerebro cognoscente que cumpliera con las condiciones del experimento. El resultado es la imposibilidad absoluta de tener un cerebro vivo y funcional apartado de toda relación con un cuerpo. El sistema cognitivo no es por tanto el cerebro, ni siquiera el sistema nervioso completo, sino todo el organismo en conjunto con un medio sobre el que pueda enactuar.

Aunque el concepto de enacción está presente en los primeros trabajos de Varela y Maturana en la década de los años 1970, es hasta la década de 1990 en que Varela (1991 y 1996), junto con su grupo de investigación, propone la neurofenomenología como el método de estudio de los procesos cognitivos. La neurofenomenología es la propuesta que pretende superar el problema de la brecha explicativa: aquella que se da entre las explicaciones objetivas de la ciencia de la mente y los reportes introspectivos de los sujetos que reportan sus propios estados mentales. De tal modo, pretende integrar la experiencia vivida en primera persona con la obtención de datos neurobiológicos en tercera persona. La interacción, dicen sus proponentes, no es por reducción, sino de carácter dialéctico manteniendo cada aspecto de la cognición su respectiva autonomía. La forma en que se pretende superar la mencionada brecha entre las perspectivas científica e introspectiva es por medio de la enacción. La tesis básica es

que la mente no es un mecanismo que tenga como principal función la de simbolizar el mundo y manipular las representaciones producidas, sino que es el medio por el que el organismo actúa sobre su mundo. No se niega que la mente sea capaz de producir representaciones. De hecho, lo hace. Lo que se niega es que esas representaciones tengan sentido mientras no sean producto de la actuación en el mundo.

Para el enactivismo, el ejercicio de los bucles de retroalimentación sensorio-motrices es lo que permite que el organismo se auto-constituya como individuo. A diferencia de las teorías representacionistas que identifican a los sistemas cognitivos en forma externa, el enactivismo afirma que los organismos no requieren ser identificados en tercera persona como individuos para actuar como tales. Es su acción lo que los auto-delimita. Al mismo tiempo, es ese límite intrínseco el que los configura en sus posibilidades de acción, de modo que se acoplan a un mundo en el que su acción se vuelve posible y con sentido. Esta idea de organismo como individuo autónomo está estrechamente ligada con el concepto de *autopoiesis* que Varela utiliza como vía de definición de los organismos vivos: son aquellos sistemas organizados capaces de llevar a cabo procesos de mantenimiento y auto-regeneración.

Si la enacción pretende ser un esquema ampliamente aceptado de explicación de la cognición en general, cabe entonces preguntarse, ¿cuáles son los procesos en que efectivamente puede ser usada la enacción como herramienta de estudio? Queda claro que los enactivistas se encuentran cómodos explicando procesos perceptuales y que implican acciones motrices dirigidas a objetivos. Se trata de procesos de bajo nivel de abstracción y que, por lo tanto, no requieren de

mucho esfuerzo para explicarlos sin recurrir a ninguna forma de representación, computación o manipulación simbólica. Lo mismo ocurre con estados propioceptivos (referentes a la percepción de la posición del propio cuerpo) y con la descripción de estados pre-reflexivos: aquellos que no requieren de ninguna forma de conciencia reflexiva o de pensamiento racional alguno. Sin embargo, no es inmediatamente claro que los principios enactivistas sean suficientes o, incluso, del todo adecuados para explicar procesos cognitivos de mayor abstracción. No obstante, al igual que los partidarios de la psicología ecológica, los enactivistas afirman que ninguno de estos procesos queda fuera de su alcance teórico: la conciencia, la producción de conceptos abstractos socialmente compartidos, el lenguaje, las matemáticas o las emociones son susceptibles de explicaciones enactivas. El problema, en todo caso, afirman los enactivistas, no está en el alcance de la teoría, sino en la articulación de las explicaciones de los diferentes niveles de estados cognitivos.⁹

Los principios enactivistas favorecen: 1. los estudios diacrónicos (los que consideran la historia de los sistemas tanto en el individuo como en la especie) sobre los sincrónicos (los que se enfocan en las relaciones de los sistemas en un momento determinado); 2. los estudios que toman a las acciones como elemento esencial; y 3. los estudios que hacen énfasis en el papel primordial de los bucles de retroalimentación sensorio-motrices. Con esto en consideración, podemos ver que existen estudios enactivistas, por ejemplo, de la adquisición del lenguaje contra el innatismo de Chomsky, tanto en el individuo como en la especie, en términos de estructuras

⁹ Stewart *et al.*, 2010

que emergen por la acción dinámica del sujeto en el medio ecológico y social.¹⁰

A partir de la afirmación de que el acoplamiento estructural permite explicar la creación de un mundo con sentido, los enactivistas infieren la posibilidad de explicar la cognición y la conducta social.¹¹ En este caso, los estudios se enfocan en atacar el individualismo tradicional de los modelos que producen las teorías clásicas de la cognición social. Para los enactivistas, la cognición social no es un proceso individual que pueda ser modelado con independencia del medio y que después deba ser considerado en forma estadística, sino que ha de ser considerado desde un inicio como un sistema complejo de interacciones que se retroalimentan entre agentes cognitivos. Para ello, al igual que para el resto de fenómenos a explicar desde la tesis de la enacción, se requieren de herramientas matemáticas de carácter dinámico.

Finalmente, hay que mencionar la vía terapéutica de aplicación de los principios de la enacción. En efecto, se ha postulado la posibilidad de una terapia enactiva para hacer consciente al sujeto de la estrecha unidad entre el cuerpo y la mente en la resolución de conflictos emocionales.¹² A partir de la idea de los bucles de retroalimentación se han producido múltiples métodos de mejoramiento cognitivo, en donde el sujeto recibe información concerniente a su desempeño en una tarea en específico. De tal manera es capaz de afinar sus bucles sensorio-motrices de modo que mejoren sus indicadores. En particular, esta técnica

¹⁰ Bottineau, 2010.

¹¹ Di Paolo *et al.*, 2010.

¹² Colombetti, 2010.

ca ha sido usada como una herramienta complementaria para que los pacientes con epilepsia puedan identificar y describir de forma más rigurosa las auras que anteceden a las crisis.¹³ Quienes proponen este uso de la enacción afirman que es posible establecer relaciones de isomorfía entre las descripciones en primera persona y los datos arrojados por los métodos de escaneo cerebral durante la crisis epiléptica.

3. Las extensiones de la mente

¿Dónde termina la mente y empieza el resto del mundo? Esta es la pregunta con la que Andy Clark y David Chalmers inician su artículo «*The extended mind*» publicado en 1998. Aunque el externalismo tiene sus antecedentes en las teorías que intentan explicar el contenido de los estados mentales (Putnam, 1975; Burge, 1979), es el texto de Clark y Chalmers el que propiamente inaugura esta nueva perspectiva sobre el estudio de la mente. La pregunta que hacen los autores plantea varias cuestiones: ¿hay un lugar en que se localice propiamente la mente?, en ese caso, ¿cuáles son sus límites?, ¿esos límites son claros e infranqueables o, por el contrario, son permeables y flexibles? Desde que Descartes ubicara a la mente en la glándula pineal, la localización espacial de la mente es todo un tema tanto filosófico como científico. Hemos visto cómo intentaban resolver el problema posturas como los fisicismos, los conductismos o los funcionalismos. Todas estas teorías asociaban la mente a una parte

¹³ Le van Quyen, 2010.

del organismo (el cerebro, un sistema o módulo neuronal) o al organismo en su conjunto (su conducta o su función). La psicología ecológica y la enacción, por su parte, hacían del cuerpo y sus interacciones en el mundo una parte constitutiva de los estados mentales, pero ninguna había asociado la mente al sistema formado por el organismo y los medios físicos externos de los que el organismo se vale para llevar a cabo sus procesos cognitivos.

Para la teoría de la mente extendida, esos medios externos no son una simple ayuda circunstancial que el organismo utiliza en su acción cognitiva, sino que ellos mismos son parte constitutiva del proceso cognitivo. Clark y Chalmers definen su postura como un *externalismo activo* consistente en afirmar que las relaciones entre el agente cognitivo y el mundo no son causales y accesorias (lo que resultaría trivial), sino constitutivas de la mente. Es decir, parte de la mente se realiza en el mundo fuera de «la piel y el cráneo» del sujeto. Dicho límite piel/cráneo es espurio para definir la cognición. El argumento para hacer esta afirmación es el siguiente: si un proceso que se lleva a cabo en el mundo (fuera del cuerpo del sujeto) realiza una función que, en caso de ser llevada a cabo dentro del cuerpo del sujeto (en «la mente») afirmaríamos que se trata de un proceso mental, entonces ese proceso es parte constitutiva del proceso cognitivo independientemente del lugar físico en el que se realice. Este argumento ha sido identificado como *argumento de la equivalencia cognitiva*.¹⁴

Bajo esta idea, si un lápiz y un cuaderno cumplen todas las mismas funciones relevantes que la memoria biológica de un sujeto, entonces no hay razón para negar que el sis-

¹⁴ Para la exposición del argumento y sus objeciones, véase Aizawa, 2014.

tema lápiz-cuaderno es parte constitutiva de la creencia del sujeto. Los proponentes de esta teoría afirman que la especie humana ha desarrollado técnicas evolutivas de descarga y soporte externo de procesos que han liberado a la mente al tiempo que la han potenciado en su desarrollo. Desde el uso de los dedos para llevar a cabo una operación, o el dibujo de esquemas y figuras para planear una acción, hasta el uso de computadoras, telescopios y cualquier tipo de herramientas de alta tecnología, todos son ejemplos de cómo la mente encuentra maneras de extender sus procesos más allá del estrecho límite del cuerpo humano.

La parte «activa» de este externalismo radica en el concepto de *acciones epistémicas*: son aquellas que el sujeto realiza en el mundo para desarrollar o, incluso, para decidir o resolver el proceso cognitivo. A diferencia de las *acciones pragmáticas* comunes que se llevan a cabo con un objetivo cognitivo decidido de antemano «dentro de la mente», las acciones epistémicas plantean o deciden dicho objetivo en su ejercicio. Ejemplos de estas acciones serían: las lluvias de ideas, el reordenamiento de las letras de *scrabble* en la tablilla para encontrar una palabra válida, la manipulación de las figuras que van apareciendo en la pantalla en el juego del Tetris o, para el caso, las complejas interacciones llevadas a cabo al interior de un acelerador de partículas sin un objetivo epistémico del todo fijado de antemano. La afirmación es que dichas acciones no son la parte mecánica de un proceso de pensamiento, sino parte del pensamiento mismo: ¡pensamos con las acciones!¹⁵

¹⁵ Aunque David Chalmers se ha alejado de este externismo epistémico en su trabajo posterior, Clark ha seguido desarrollado la tesis en sus libros recientes (Clark, 2004 y 2016).

A favor de la tesis de la mente extendida parecen estar los avances en materia de implantes neuro-tecnológicos y extensiones corporales. Así como llevamos la mente hacia el mundo por medio de instrumentos que nos permiten elaborar operaciones y recordar información, de manera simétrica parece que podemos hoy llevar el mundo hacia la mente por medio de implantes que permiten formas de percepción visual en invidentes estimulando zonas específicas de la lengua para aprender a reconocer objetos en el espacio circundante del sujeto, o por medio de sensores que son capaces de «leer» señales cerebrales para mover miembros artificiales. Es de esperar que tecnologías como estas sean cada vez más avanzadas y accesibles de modo que el reto a la confinación teórica de la mente dentro del cráneo sea cada vez mayor.¹⁶

Una posible objeción radica en afirmar que, no obstante, hay ciertos estados cognitivos que «claramente» se realizan por completo «dentro de la mente» de modo tal que la acción que producen sólo es el producto conductual del proceso terminado. Y, en efecto, parece que eso sucede con frecuencia. Los extensionistas podrían proceder de dos maneras: 1) afirmando que en realidad todo proceso cognitivo tiene partes de él que bien pueden ser «descargadas» en el medio, en cuyo caso resulta irrelevante si se realizan o no «dentro de la mente», o 2) pueden sostener una postura más adelgazada, afirmando que si bien no todo proceso cognitivo está determinado externamente, algunos sí lo están, en cuyo caso

¹⁶ Sobre el tema de las prótesis tecnológicas y las interacciones cerebro-computadora en general se puede ver el libro de Clark referido en la nota anterior (Clark, 2004) y la muy reciente y completa compilación: *Brain-computer interfaces handbook: technological and theoretical advances* (Nam *et al.*, 2018).

la mente se extiende en el mundo. Lo que el extensionista habrá conseguido con esta segunda opción será, de cualquier modo, atacar la idea de los estados cognitivos dentro de la mente como clase natural, es decir, como un conjunto de entidades que tienen una propiedad en exclusiva que los define claramente (su cualidad de «estar dentro» del cráneo, por ejemplo). En este caso, la teoría de la mente extendida podría afirmar que los procesos cognitivos cotidianos son extendidos, pero no así, quizá, los procesos cognitivos involucrados en el quehacer científico.

La pregunta final sería, si ya reconocimos que hay ciertos procesos cognitivos que se extienden más allá del cuerpo, ¿en dónde está el límite? Clark y Chalmers llegan a afirmar: ¿y por qué poner un límite? El lenguaje, a final de cuentas, es una herramienta que evolucionó justamente como forma de extensión del pensamiento. Y es el lenguaje el que posibilita la vida social, de modo que, incluso, los medios externos de mis procesos cognitivos pueden ser los estados mentales de otras personas, valiéndonos de ellas y de sus propios procesos para completar los míos. Y es entonces cuando tenemos que comenzar a considerar los temas de la intersubjetividad y la cognición social.

4. Interludio: neuronas y el espejo de la intersubjetividad

Imagina la siguiente escena: el protagonista de la película de acción que estás viendo cuelga desesperado de un risco a punto de caer en un profundo abismo agarrado tan solo de una mano. La toma permite ver el rostro desencajado, así

como la mano en un gesto contraído. El peso es demasiado para los dedos que comienzan a ceder uno a uno. El sudor colabora a que la mano pierda agarre. La caída parece inminente. La cámara hace *close up* alternativo tanto al rostro como a la mano...

A pesar de que la narración pueda ser perfectible y que no tenga la viveza de la imagen, seguro pudiste adivinar emociones en el personaje (aunque no se mencionara ninguna explícitamente): angustia, terror, desesperación. Quizá sentiste también un cosquilleo incómodo en los dedos al tiempo que leías cómo la mano hacía un esfuerzo desesperado por agarrarse del risco. Es la misma sensación incómoda que se tiene al ver el momento en que un deportista se fractura una pierna o en que la bailarina del espectáculo que estás presenciando realiza una contorsión imposible para ti. ¿Cómo sucede esto?, ¿cómo es que la observación (o la narración) de lo que sucede en un cuerpo ajeno desencadena sensaciones en mi propio cuerpo?, ¿cuál es el mecanismo neural subyacente a semejantes reacciones?

En 1992, mientras se encontraban realizando un estudio con un objetivo diferente, un grupo de neurocientíficos en Parma, Italia, descubrió un conjunto de neuronas en la zona F5 del cerebro de los *macacos nemestrinos* que presentaban características que no habían sido previamente registradas. A estas neuronas las bautizaron como *neuronas espejo* (NE).¹⁷ Se sabía con anticipación de la existencia de cierto grupo neuronal, tanto en humanos como en primates, que tienen dos modos de activación: ante estímulos visuales y ante actos motrices. Estos dos modos de activación resultan congruen-

¹⁷ Di Pellegrino *et al.*, 1992.

tes, es decir, las neuronas de este grupo se activan tanto al ver un objeto determinado como al actuar sobre él. Esto, por sí mismo, ya es sorprendente, pues desafía la idea especulativa de que los sistemas encargados de la percepción y de la acción son distintos.¹⁸ La sugerencia de la existencia de estas neuronas es muy temprana en la historia de la neurociencia: aproximadamente hacia la década de 1930; es por eso que se les nombró como *neuronas canónicas* (NC). Las NC están muy especializadas, de modo que, en conjunto, brindan el amplio espectro de movimientos sutiles que el mono (y el ser humano) pueden hacer con la mano: algunas neuronas se encargan de agarres con toda la mano, otras se activan al tomar los objetos con dos dedos y otras más se encargan de los movimientos para sostener un objeto con un solo dedo en forma de gancho.

Lo interesante es que, mientras los neurocientíficos italianos estudiaban estas neuronas en los macacos, encontraron que algunas de ellas presentaban características ligera y significativamente distintas. Las NE también son neuronas bimodales que se activan por igual con estímulos visuales y con acciones motrices. La diferencia está en el tipo de estímulos visuales que las excitan. A diferencia de las NC, las NE no reaccionan ante la visión de objetos, sino ante acciones ejecutadas por otro sujeto (el investigador u otro mono). De

¹⁸ La existencia de estas neuronas bimodales puede ser vista como un argumento para reforzar las tesis tanto de la psicología ecológica como de la enacción. En efecto, parece que la activación visual obedece a las posibilidades de actuar sobre el objeto; esta es la idea básica de los *affordances* de Gibson. Del mismo modo, esto comprobaría empíricamente la idea de que el conocimiento es un bucle de retroalimentación entre la percepción del medio y la acción del sujeto.

las 184 neuronas que estudiaron en ese primer experimento, casi 90 presentaron esta curiosa bimodalidad: acción observada-acción ejecutada. Las acciones que se presentaron ante los macacos fueron variadas: poner comida sobre una superficie, recogerla, dar un objeto a otro sujeto, manipular cualquier objeto, hacer gestos con contenido emotivo (amenazar, levantar los brazos en símbolo de alegría, saludar). En todos estos casos, las NE correspondientes se activaron por igual cuando el macaco realizó las mismas acciones.

Hay un par de casos en los que es necesario poner atención. En primer lugar, ninguna de las NE estudiadas respondió ante gestos observados sin un objetivo claro de por medio: agitar las manos sin comida, mover la mano lejos del objeto, tomar el objeto con una herramienta. En segundo lugar, hubo algunas pocas neuronas que presentaban una bimodalidad no perfectamente congruente, sino sólo *lógicamente* congruente. Es decir, la misma neurona disparaba ante un acto lógicamente relacionado con aquel que ejecutaba el macaco: por ejemplo, se activaba al ver al experimentador colocando la comida sobre el plato al igual que al tomarla por parte del macaco. De todo esto, los investigadores concluyeron que las NE reaccionan solo ante *acciones significativas*, seleccionando los movimientos que permiten comprender las intenciones de otros individuos permitiendo el establecimiento de relaciones interpersonales.

Desde aquel primer hallazgo han pasado ya 30 años. Desde entonces, el estudio de las NE se ha consolidado como uno de los más fecundos en la neurociencia cognitiva contemporánea. Posteriores investigaciones descubrieron múltiples sistemas con cualidades bimodales espejo. Se han observado neuronas con activaciones no solo para acciones con

las manos, sino también con la boca.¹⁹ De la misma forma, se observaron sistemas espejo en regiones cerebrales como la ínsula y el cíngulo anterior para emociones básicas como el asco, el enojo o el miedo. Se han descubierto sistemas similares tanto en monos titíes como en aves. En estas últimas, las NE implicadas reaccionan ante estímulos sonoros. Esto ha permitido abrir una línea de investigación sobre la participación de los sistemas espejo en el aprendizaje: para cantar, en el caso de las aves, o para aprender a seleccionar rutinas motrices significativas en el caso de los monos y los humanos. A partir de la resonancia sonora se puede entender que las NE tienen un importante papel en la investigación sobre el desarrollo del lenguaje humano tanto en la especie como en el individuo.²⁰ Según los propios neurocientíficos involucrados en la investigación de los sistemas espejo, se esperan avances futuros en el descubrimiento de los mecanismos neuroquímicos y moleculares subyacentes a estos procesos de resonancia. Toda esta información podrá ser de utilidad no solo para la comprensión de las capacidades humanas de interacción social, sino para el estudio de desórdenes neurológicos y psiquiátricos como el autismo, la esquizofrenia o algunos síndromes delirantes.

Como se puede apreciar, el impacto fuera de la neurociencia, en disciplinas como la psicología, la sociología, la etología, la lingüística o la filosofía, ha sido de gran repercusión. El hallazgo puso en el centro del debate la cuestión de

¹⁹ La boca y las manos son dos de las partes del cuerpo más representadas en la corteza cerebral, tanto en la zona motriz, como en la sensorial.

²⁰ Al respecto, es interesante notar que el área F5 de los macacos corresponde, más o menos, al área de Broca en el cerebro humano, la cual es la encargada de una parte del procesamiento del lenguaje.

cómo comprendemos las acciones e intenciones ajenas. Para Di Pellegrino, Rizzolatti, Sinigaglia y compañía,²¹ la función de las NE no es propiamente la preparación para la acción ni la imitación, sino el «reconocimiento» y la «comprensión» de los actos ajenos. De tal modo, gracias a la acción de estas neuronas bimodales, se puede explicar, afirman los investigadores, cómo es que «percibimos significados» y «comprendemos acciones» gracias a un «vocabulario motor» en un nivel neuronal. A partir de ahí, se han aventurado hipótesis sobre el origen y la función evolutiva de la cultura tomando la resonancia neuronal como centro de la explicación. De igual manera, se ha intentado ver en esta bimodalidad neural la solución de los debates filosóficos sobre las mentes ajenas cuyo origen se remonta, cuando menos, hasta la filosofía cartesiana.

Sin embargo, es necesario tomar un poco de distancia racional ante la evidencia. No todos los neurocientíficos y filósofos están tan entusiasmados con el hallazgo de las NE o, cuando menos, no todos se permiten derivar las mismas consecuencias de amplio alcance sobre tales hallazgos.²² A final de cuentas, explicar el surgimiento de la cultura a partir de un conjunto neuronal, por importante que éste sea, implica una gran cantidad de pasos en la cadena de inferencias que

²¹ Rizzolatti y Sinigaglia, 2006.

²² Se pueden revisar algunas de las objeciones a las inferencias teóricas sobre las NE en el más reciente libro de Patricia Churchland (2016). Pero si se trata de criticar la actitud entusiasta de muchos científicos y filósofos sobre este tipo de neuronas, es necesario consultar el libro de Hickok (2014) en donde el autor intenta desmontar muchos de los mitos, vaguedades y errores que abundan entre los partidarios de las NE como fundamento último de la comprensión social.

pueden hacer fallar el argumento. Por muy consolidada que esté la evidencia sobre las NE en estos últimos 30 años, es necesario hacer algunas precisiones. De inicio, hay que indicar la asimetría en la forma de obtener los datos neurales entre los humanos y los monos. En éstos últimos, es posible hacer trepanaciones craneales para registrar la actividad de neuronas individuales usando microelectrodos de tungsteno. Para ello, la cabeza del mono se encuentra inmovilizada mirando a un punto en específico. En los humanos, los estudios de ese tipo resultan imposibles por violar las normas éticas de la experimentación.²³ Entonces los datos sobre los sistemas espejo de nuestros cerebros los obtenemos por medio de métodos no invasivos (aquellos que no requieren abrir el cráneo),²⁴ como: resonancias magnéticas funcionales (fMRI), tomografías por emisión de positrones (PET) o estimulación magnética transcraneal (EMT). Cada una de estas técnicas tiene sus restricciones y sirve para obtener cierto tipo de datos que no necesariamente es evidencia directa de la activación neuronal que quisiéramos estudiar. A partir de esta asimetría metodológica, hay quien afirma que no hay evidencia concluyente de la existencia de NE en humanos.²⁵

²³ Curiosamente, consideramos a los macacos como suficientemente semejantes a nosotros para poder extender las conclusiones de la experimentación hacia nuestro propio cerebro, pero también preferimos pensarlos como suficientemente lejanos para no asignarles los derechos éticos que prohíban experimentar con ellos de la forma en que lo hacemos.

²⁴ Casos excepcionales son aquellos en que se realizan cirugías cerebrales. En esas condiciones es posible estudiar el cerebro en forma directa pues hay un objetivo médico de por medio para abrir el cráneo. No obstante, las condiciones experimentales en estos casos se encuentran muy restringidas y subordinadas al objetivo por el cual se lleva a cabo la cirugía.

²⁵ Lingnau A, Gesierich B, Caramazza A., 2009.

Desde el punto de vista netamente filosófico, también hay que hacer una apelación a la prudencia. Como vimos desde el primer capítulo, la TCRM tiende a ser la forma estándar en que tanto filósofos como científicos interpretan los propios hallazgos empíricos. Parte de este marco conceptual implica adjudicar cualidades mentales a mecanismos artificiales (lo que conocemos como IA fuerte) o a las partes de la unidad que es una persona. En otras palabras, se tiende a afirmar que el cerebro en sí mismo «piensa», «decide» o «siente» en forma no consciente para el individuo. Siguiendo una intuición original de Wittgenstein, hay filósofos para quienes esta práctica implica un claro error de razonamiento: la *falacia mereológica*.²⁶ Esta falacia consiste en adjudicar a una parte del sistema (el cerebro, o un grupo neuronal, en este caso) propiedades que sólo pueden ser adjudicadas al sistema en su totalidad: pensar, interpretar, comprender o decidir.

Cuando los científicos del grupo de Parma extraen las conclusiones de sus experimentos hacen afirmaciones que *intencionalizan* a un grupo neuronal, asignando predicados mentales a una parte muy restringida del sistema nervioso central. Se habla de comprensión de acciones, de significados y de intenciones en circuitos neuronales. De la misma manera, se habla de vocabularios motores cifrados en dichos sistemas espejo. Siempre es posible justificar tales expresiones diciendo que se trata de una forma analógica de expresión que sólo intenta hacer comprensible lo que sucede a nivel cerebral. En realidad, podrían decir los investigadores, nadie afirma que el cerebro *literalmente* toma decisiones o tiene emociones. No obstante, la mayoría de las veces se trata de algo más que

²⁶ Bennett y Hacker, 2003.

«una forma de hablar». Tales expresiones tienen consecuencias tanto en las inferencias teóricas como en la dirección de los experimentos y el desarrollo de la investigación futura. Desde el punto de vista filosófico, por ejemplo, tal intencionalización implica afirmar que el sujeto, en realidad, es el cerebro o una parte de él. Lo cual, lejos de resolver el problema, nos abre la puerta hacia un regreso al infinito. Y en todo caso, si el cerebro tiene sus percepciones y decisiones propias, cabría preguntarse, ¿cómo se relacionan esos estados mentales con los del sujeto?, en última instancia, ¿cómo se relacionan *ambos sujetos* (el individuo en su conjunto y el subpersonal)? En cuanto al aspecto empírico, afirmar cualidades mentales al cerebro nos lanza a la búsqueda de un lugar, un subsistema o una función en específico que sea la responsable de la «mentalidad». Es decir, estaríamos buscando una especie de homúnculo o fantasma en la máquina que, según todo lo que sabemos, parece dirigirnos hacia una búsqueda estéril.

No obstante, hoy podemos afirmar que hay suficiente evidencia para pensar que efectivamente existen sistemas espejo en los seres humanos. Podemos afirmar también que dichos sistemas juegan un papel importante en nuestra comprensión de la conducta ajena y en el establecimiento de los lazos sociales en general. De qué forma sucede esto y en qué colaboran estos hallazgos a los debates filosóficos sobre la comprensión social, lo veremos a continuación.

5. Comprensión, cultura y empatía

Ana tiene un juguete. Ana entra a su cuarto y guarda el juguete dentro de la caja A, después sale del cuarto. Sin que

Ana lo sepa, entra su hermano Juan al cuarto y cambia el juguete de la caja A a la caja B. Cuando Ana regresa a buscar su juguete, ¿en dónde lo buscará? Si tienes más de cinco años, seguramente contestarás que Ana buscará en la caja A, que es donde dejó el juguete que busca. Los niños menores de cinco años tienden a responder que Ana buscará el juguete dentro de la caja B, es decir, tienden a asignarle al personaje de la historia el conocimiento que ellos mismos tienen. A este experimento se le conoce como el *test de la creencia falsa*. El niño de cuatro años no pasa el test pues no es capaz de asignar la creencia falsa de que «el juguete está en la caja A» al personaje. ¿Qué es exactamente lo que separa esta línea entre los cuatro y los cinco años de edad?, ¿existe en efecto tal línea divisoria? Y, en todo caso, ¿qué dice esto sobre la forma en que desarrollamos nuestras habilidades de interacción social?

Al final del apartado sobre la mente extendida abrimos la posibilidad de considerar a la cognición en una perspectiva social. El interludio sobre las NE nos ha dado las herramientas para entender el debate actual sobre la realización neural que pueda explicar nuestras cualidades de interacción social y la comprensión de la conducta de nuestros semejantes. No obstante, como sucede con cualquier concepto en filosofía, el de «cognición social» no es un concepto libre de ambigüedades. En efecto, cuando hablamos de cognición social podemos entender alguno de los siguientes proyectos:

1. El proyecto conocido como *epistemología social* que busca investigar las condiciones, elementos y efectos sociales de las acciones epistémicas llevadas a cabo en una sociedad históricamente determinada. Para los trabajos que se inscriben dentro de esta corriente teó-

rica, la epistemología tradicional ha abordado en forma incorrecta el problema del conocimiento desde un inicio al concebirlo como una empresa eminentemente individual cuyo resultado, si acaso, tiene consecuencias sociales en forma contingente. En cambio, para la epistemología social no puede comprenderse ninguna acción cognitiva si no es dentro de un sistema de relaciones e instituciones sociales dado de antemano a donde el individuo llega a insertarse desde un inicio. Por lo tanto, vista desde esta perspectiva, la epistemología social no es una teoría epistemológica más, sino una forma de reivindicación del conocimiento desde su estructura esencial. Antecedentes de este proyecto se encuentran en los trabajos de Kuhn (1962), Foucault (1961, 1966, 1969), y en los teóricos de la sociología del conocimiento (Bloor, 1976; Barnes, 1974; Latour y Woolgar, 1979). El desarrollo contemporáneo de la epistemología social se encuentra en autores como Alvin Goldman (Goldman, 1987, Gallese y Goldman, 1998) y Philip Kitcher (1993). Los debates sobre los que gira el trabajo de este proyecto acostumbran a relacionarse con la evidencia social, los testimonios y el desacuerdo en casos de controversia jurídica; la indagación sobre si tiene sentido hablar de grupos sociales como agentes de conocimiento; y la relación entre argumentación, ciencia, derecho y democracia. Pero, sobre todo, una vertiente importante de la epistemología social ha sido desarrollada en el ámbito pedagógico en donde se han implementado modelos educativos a partir de una concepción amplia y enriquecida de los

procesos cognitivos que integran elementos afectivos y de interacción social (Barton y Garvis, 2019).

2. El proyecto que teoriza sobre el concepto de *intencionalidad colectiva*, bajo el cual se pretende argumentar que existen relaciones intencionales propias de los grupos sociales que no pueden ser reducidas a las formas de intencionalidad mental de los estados de los sujetos individuales. El principal defensor de esta idea es John Searle, quien afirma que hay una estructura lógica clara y formalizable detrás de acciones colectivas que involucran la planeación y la cooperación. Dichas acciones son pre-lingüísticas y según Searle, son el fundamento de toda la estructura social. En una acción cooperativa, los individuos pueden llevar a cabo tareas individuales diferentes entre sí, con contenidos distintos y, sin embargo, compartir el mismo contenido intencional: el mismo objetivo consciente y voluntario. De esta forma se pueden explicar desde las promesas y los compromisos, hasta los deportes, las ejecuciones musicales de un grupo, las empresas, las instituciones y los gobiernos.²⁷
3. El proyecto conocido como *teoría de la mente*, el cual pretende explicar nuestra habilidad para comprender, describir y predecir los estados mentales y la conducta de los otros sujetos.

Por tratarse del tema con mayor relación con las propuestas corporizadas que estamos revisando en este capítulo, a con-

²⁷ Sobre intencionalidad colectiva, véase Searle, 2014.

tinuación, revisaremos con detalle el proyecto de la teoría de la mente.

Cuando Descartes, al asomarse por su ventana, afirma que no puede saber con total certeza si aquello que ve pasar es una persona o un simple autómatas inaugura el *problema de las otras mentes*.²⁸ Para la perspectiva racionalista moderna, y para la TCRM, saber si los sujetos con los que me encuentro tienen una mente es, en efecto, un problema. Se asume que la única forma de constatar la existencia de la mente con total certeza (apodíctica) es por vía introspectiva, es decir, en primera persona. De tal manera, uno sólo podría estar seguro de la existencia de su propia mente. La afirmación de que cualquier otra persona posee una mente, o estados mentales similares a los míos, sería sólo una inferencia más o menos probable. No obstante, hay que reconocer que los seres humanos somos sujetos altamente eficaces en el día a día en el reconocimiento, interpretación y predicción de la conducta ajena. Somos capaces, sin apenas notarlo, de explicar la conducta de otras personas por medio de sus creencias y motivaciones. De igual manera, sabemos con bastante certeza cuál será la reacción de una persona ante una noticia de acuerdo al grado de conocimiento que tengamos sobre ella. Muchas veces, incluso sin conocer a la persona en cuestión, somos capaces de saber con precisión el estado mental general de una persona sólo al ver sus gestos: tristeza, dolor, asco, ira, felicidad, etcétera. Para todo esto apenas y requerimos de estados de nivel cognitivo o reflexivo superior; parecieran, en realidad, procesos instantáneos de percepción directa de los

²⁸ Véase la segunda *Meditación* (Descartes, 2014).

estados mentales correspondientes.²⁹ A esta habilidad social se le ha nombrado a partir de un anglicismo: lectura mental (*mind reading*). La pregunta, entonces es, ¿cómo es posible esa eficacia social? Y, en todo caso, ¿cuál es el proceso que explica dicha «lectura mental»?

La filosofía de la mente de corte cognitivista dispone de un par de posibles respuestas excluyentes entre sí. Por un lado, está la *teoría-teoría* (TT) y, por el otro, tenemos la *teoría de la simulación* (TS). Para la TT la respuesta a la pregunta sobre nuestra eficacia social está en la posesión de una teoría sobre el funcionamiento de las mentes en general que el sujeto posee y que se refina progresivamente a medida que la persona crece y se socializa. Dicha teoría constaría, al igual que una teoría científica, de conceptos, leyes y presuposiciones sobre la forma en que funcionan las mentes. En el mundo científico disponemos de teorías semejantes para comprender el funcionamiento y estructura de la mente. A la ciencia encargada de dichas elaboraciones teóricas la llamamos psicología. La diferencia es que la teoría a la que nos referimos en la TT no es expresa, nadie la enseña al niño en forma académica y se desarrolla de manera natural desde el nacimiento. Pero, no obstante, seguiría siendo una forma de psicología. Es por eso que se le ha dado en llamar: psicología popular (*folk psychology*).³⁰ De esta manera, la conducta observada en los otros sujetos es la evidencia que el individuo usa como insumo para inferir y comprobar sus conjeturas

²⁹ El fenomenólogo Max Scheler decía que, desde la perspectiva de primera persona, no sucede que vea lágrimas en un rostro ajeno e infiera la tristeza, sino que «veo tristeza en las lágrimas» (Scheler, 2007).

³⁰ En algunas ocasiones el término se ha traducido también como «psicología del sentido común» o «psicología ingenua».

sobre los estados mentales de las otras personas. Para los proponentes de la TT, el fallo en el test de la creencia falsa de los niños menores a cinco años de edad se explica porque todavía no poseen ese concepto en particular (el de «creencia falsa»), de modo que no son capaces de prever acertadamente la conducta de la persona en el escenario supuesto.³¹

Por su parte, la TS rechaza que se trate de una inferencia teórica a partir de la conducta observada. Los niños no llevan a cabo procesos mentales cuasi-científicos para relacionarse socialmente. En realidad, afirman los teóricos de la simulación, el proceso es el siguiente: la persona observa las circunstancias en que se encuentra el otro, a partir de ahí, «simula» un escenario hipotético en donde él mismo se encontrara en dicha circunstancia. De tal modo, los estados mentales que la persona identificara en sí misma dentro de tal simulación, los puede extrapolar analógicamente a la persona cuyos estados mentales pretende comprender. Es como si corriera el programa *off-line*, para después probarlo *on-line* explicando la conducta realmente observada en el otro sujeto. Así, según los simulacionistas, lo que los niños menores a cinco años son incapaces de hacer cuando fallan el test de la creencia falsa es ubicarse a sí mismos como teniendo una creencia distinta (falsa) a la que en realidad tienen (verdadera). El problema está en su incapacidad de simulación de escenarios hipotéticos (contrafácticos, dicen los filósofos) y no en la carencia del concepto adecuado para explicar la conducta observada.³²

³¹ Sobre la TT, véase Gopnik y Meltzoff, 1997 y Gopnik y Wellman, 1994.

³² Sobre la TS, véase Gordon, 1986; Goldman, 2006; Goldman y Jordan, 2013.

¿Quién tiene la razón?, ¿es posible que el mismo experimento, con los mismos resultados pueda servir al mismo tiempo como evidencia empírica a favor de dos teorías mutuamente incompatibles? Aquí hay que recordar que ambas respuestas, tanto la TT como la TS son teorías de corte cognitivista. Desde el punto de vista de la corporización, ambas opciones son erróneas desde su planteamiento original. Si esto es cierto, entonces la disputa entre TT y TS es un pseudoproblema. El error radica en asumir que para explicar nuestra habilidad social se requiere comprender o una teoría (en el caso de la TT) o un razonamiento analógico que no deja de ser teórico (en el caso de la TS). Pero, en realidad, ni la teoría ni el razonamiento son necesarios. La cognición corporizada pretende proponer una explicación deflacionaria: una que no requiera de elementos especulativos como teorías o leyes, ni de procesos de alta jerarquía cognitiva como inferencias, analogías o simulaciones. A fin de cuentas, salvo en casos muy específicos y anómalos, parece compatible con nuestra vivencia cotidiana que no llevemos a cabo procesos de este tipo cuando interactuamos socialmente.

La teoría de la *cognición social corporizada* (CSC) propone que nuestras interacciones no se explican por medio de la adscripción de estados mentales a los otros sujetos, sino por medio de una intersubjetividad primaria, pre-teórica y no conceptual que subyace a las habilidades sociales y cognitivas de mayor complejidad. Según autores como Gallagher (2005) y Hutto (2008), dicha subjetividad primaria se evidencia desde la primera infancia en distintas conductas como en la habilidad de imitación facial, la propiocepción, el rastreo de la mirada de los otros para adivinar sus pensamientos e intenciones o en la comprensión de emociones en los gestos

corporales. A partir de esta primera forma de intersubjetividad, se construye una segunda, más compleja y específica, que estaría basada, principalmente, en la comunicación lingüística. Según la CSC, nuestras relaciones sociales cotidianas no están determinadas principalmente por una lectura de la mente tal como lo proponían las respuestas cognitivistas. Esto no significa que no seamos capaces de realizar tal lectura o que, en algunos casos recurramos de hecho a ella. En ocasiones, es necesario un proceso reflexivo y detenido para intentar comprender la conducta de los otros. No siempre nos son evidentes las intenciones o las emociones de los demás: conductas excéntricas o inesperadas tienden a confundirnos y a hacer que nos preguntemos explícitamente por lo que la persona en cuestión está pensando. No obstante, afirman los teóricos de la CSC, esto sucede solo cuando las prácticas corporizadas tradicionales dentro de una comunidad determinada son ambiguas. En estos casos, nuestros modos primarios de cognición social fracasan. Y entonces nos vemos obligados a recurrir a estrategias secundarias de carácter lingüístico: preguntar explícitamente a la persona cuya conducta nos desconcierta.

A pesar de que los niños pequeños no son capaces de llevar a cabo esa lectura de la mente, tienen interacciones exitosas con su círculo social. Eso explica que desde edades pre-lingüísticas los bebés pueden ser capaces de exhibir una comprensión consistente de las emociones y creencias de sus semejantes. Pero esto explica también que los niños menores a los cinco años fracasen en el test de la creencia falsa. Para la CSC el fracaso se debe a que el test está elaborado sobre la hipótesis de la teoría de la mente. El resultado es un diseño experimental que suprime las circunstancias corporales y

contextuales a las que los niños están habituados en condiciones normales. Si esto es cierto, entonces en realidad el test de la creencia falsa es un mal criterio para evaluar la interacción social en la infancia, pues la facultad que presuntamente mide no sería ni necesaria, ni suficiente para explicar el desarrollo social de las personas. Para demostrar esto, se han desarrollado experimentos similares al de la creencia falsa, pero con un diseño que no requiere del lenguaje. De esta manera se ha podido observar que en edades cercanas a los 15 meses (muy inferiores a las del test tradicional) los niños son capaces de prever exitosamente la conducta de Ana una vez que regrese a buscar su juguete.

Las teorías de la CSC han dado el salto, de la explicación de la interacción social individual, a la de los fundamentos de la cultura. En disciplinas como la antropología o la sociología hay un amplio espectro de conceptos de cultura. Para las propuestas corporizadas, el problema radica en que los más influyentes de esos conceptos ponen el énfasis en conjuntos de reglas, normas, tradiciones y representaciones simbólicas. Es decir, la cultura es definida en términos prescriptivos que presuntamente regulan la conducta de los individuos. No obstante, parece haber una constante brecha entre la perspectiva normativa de la cultura y la descripción de la forma en que en efecto proceden los individuos dentro de un grupo social. Esa incompatibilidad entre las normas proposicionales y la conducta manifiesta puede ser superada partiendo de una premisa corporizada. Recordemos que para las teorías corporizadas el elemento clave es la consideración de un cuerpo en su contexto. La conducta, por más que esté dirigida por reglas y normas abstractas, es resultado de contextos específicos: condiciones sociales,

estados corporales y los objetivos reales del individuo. La conducta se regula constantemente mediante interacciones corporales. Entonces, una propuesta corporizada definiría la cultura como una «puesta a punto» de rutinas sensorio-motrices de interacción social.³³

Esas rutinas sensorio-motrices serían la base de la intersubjetividad primaria que mencionan Gallagher y Hutto. Se trataría de patrones de interacción que funcionan por bucles de retroalimentación tal como lo indica la tesis de la enacción. Esos patrones están presentes desde el nacimiento y son refinados gracias a la plasticidad del sistema nervioso del bebé y al aprendizaje. En todo este proceso, las neuronas espejo intervienen permitiendo la comprensión motriz de las acciones observadas en el entorno familiar. Así, el niño se va convirtiendo progresivamente en un individuo socializado dentro de su propio grupo cercano. Esto sucede no por el aprendizaje abstracto de normas a seguir, sino por una modelación contextual, constante y progresiva del comportamiento. Con el tiempo, estas rutinas adquieren un valor normativo, es decir, se convierten en normas expresas de comportamiento en las cuales se basan los juicios que el individuo hace del comportamiento propio y ajeno. Esta sería la base de la intersubjetividad secundaria. A diferencia de la perspectiva tradicional normativa, la corporización hace depender las normas abstractas de la conducta contextual y no a la inversa.

Que las rutinas sensorio-motrices sean la forma en que un individuo se socializa, explicaría el fuerte apego que cada quien siente hacia la conducta de su grupo cercano y la

33 Véase Soliman, Gibson y Glenberg, 2013.

gran variedad de discriminaciones que los humanos somos capaces de hacer con base en distinciones mínimas entre subgrupos al interior de una sociedad. Para la CSC la discriminación, los prejuicios y la distinción entre «nosotros» y «los otros» está arraigada en nuestras formas corporizadas e inconscientes de conducta. Tendemos a socializar más y, por lo tanto, a mostrarnos más solidarios, con aquellos cuyo repertorio de modos corporales se parece más al nuestro. Así, la corporización brinda una posible forma de explicación y de exploración empírica de la solidaridad, la indiferencia, la discriminación, los estereotipos, la violencia o los prejuicios entre comunidades.³⁴

En este punto resulta necesario poner atención en las teorías pedagógicas de la cognición social que conciben la formación de los individuos ya no como un conjunto de facultades cognitivo-rationales, sino de forma que se integran las variables corporizadas y afectivas a partir del concepto de empatía. Si se entiende a la mente como un dispositivo de manipulación sintáctica de representaciones, entonces de ahí se deriva una pedagogía basada en desarrollar las habilidades algorítmicas racionales de los individuos con abstracción del resto de sus condiciones mentales y sociales. El resultado de esto es una enseñanza sin vínculo ni significado con la vida real de los sujetos y que, por lo tanto, resulta dificultosa y poco útil e interesante para los alumnos. Para las teorías de esta pedagogía social, los sujetos se hacen tales en el proceso socio-cultural de desarrollarse en entornos de interacción con otros individuos. Las consecuencias epistémicas de esto

³⁴ Sobre el estudio corporizado de los prejuicios, véase Smith y Semin, 2007.

son que el sujeto se hace consciente de sus propias capacidades y destrezas a partir de verlas reflejadas en la conducta ajena. Para ello, es necesario que a la base se encuentre una forma muy primitiva de interacción social basada en el reconocimiento de los otros como semejantes de forma tal que se genere empatía. La empatía, a su vez, será el estado mental propio gracias al cual es posible establecer comparaciones y diferencias tanto en forma física como emocional, conductual y, eventualmente, cognitiva.³⁵

Como se puede ver, para la pedagogía corporizada la empatía y la interacción social son presupuestos epistémicos que determinan tanto la forma de concebir al sujeto como de comprender la forma en que dicho sujeto se hace eficaz en el medio: a partir del reconocimiento del otro como un análogo (en parte semejante, en parte diferente). Sobre esta base, es posible el desarrollo de métodos y técnicas pedagógicas que deriven en sistemas curriculares que tengan en el centro no al individuo, sino a la interacción social.³⁶

Para la pedagogía social no basta con dejar de concebir al sujeto como un artefacto cognitivo aislado. El reconocimiento intencional de los otros en tanto sujetos y las interacciones sociales que son vitales durante las primeras etapas del desarrollo (y que intentan ser puestas en evidencia en el test de la creencia falsa) se integran con los aspectos emocionales para producir conducta eficaz e inteligente en el medio. Desde hace algunos años se ha constatado empíricamente la relevancia de las emociones en el desarrollo cognitivo de los

³⁵ Sobre el tema de la empatía, véase la compilación de Maibom (2017).

³⁶ Al respecto, véase: Demetriou (2018), Barton y Garvis (2019) y García-Ochoa y McDonald (2021).

individuos (Damasio, 1999; 2016; 2019). Las emociones son el vehículo por el cual somos capaces de valorar los estímulos y las situaciones de modo que seamos eficaces en nuestras acciones cotidianas. Sin tales valoraciones, el sujeto no será capaz de tomar decisiones enteramente racionales (Kahneman, 2015). Una pedagogía integral, por tanto, tendrá que asumir el compromiso por el desarrollo de los aspectos tanto cognitivos como emocionales del estudiante. Cuando la relación entre facultades racionales y emotivas no se encuentra correctamente equilibrada, entonces nos enfrentamos con patologías del aprendizaje y la conducta. A algunas de estas patologías dedicaremos el siguiente capítulo.

III. LA MENTE DEFICIENTE

Imagina que estás frente a una herramienta compleja cuyo funcionamiento desconoces. Lo único que sabes de ella es la función que realiza. El resultado de su actividad es la realización de una tarea clara y manifiesta: una operación compleja, por ejemplo. No sabes cuál sea su estructura interna, ni la forma y función de sus piezas, ni tan siquiera, el material del que está compuesta. ¿Qué puedes hacer para entenderla por completo? Hay varias posibilidades: abrirla y echar un ojo a su interior; intentar inferir el funcionamiento interno a partir de su acción visible; crear otra máquina que intente replicar el funcionamiento de la máquina desconocida de modo que lo que infieras de la segunda podrá ser extendido por analogía respecto de la primera; o puedes ver los casos en que la máquina deja de funcionar para inferir por vía negativa su funcionamiento normal.

En realidad, todas estas son las vías que tenemos disponibles para el estudio de la mente: la observación conductual, el estudio de la anatomía y de la fisiología del sistema nervioso, la elaboración de modelos computacionales o el estudio de los trastornos mentales. A cada una de estas vías

de estudio corresponden diferentes disciplinas científicas: la psicología, las neurociencias, la IA y la psicopatología. Pero a diferencia de la herramienta compleja cuyo funcionamiento desconocemos, las mentes son propiedades de organismos vivos, de modo que tenemos también otras vías de investigación: podemos estudiar la historia evolutiva de las especies que exhiben mentalidad hoy en día, podemos poner atención en la forma en que los nuevos miembros de dichas especies se desarrollan y llegan a la madurez o, incluso, podemos investigar a especies actuales semejantes a la que nos interesa para extraer conclusiones comparativas. Así, tenemos que también nos son de ayuda disciplinas como la biología evolutiva, la psicología del desarrollo, la etología, o la psicología evolucionista. Cada una de estas vías de investigación arroja información útil pero parcial: estructura, desarrollo, función. La información obtenida por los investigadores en estas áreas con frecuencia debe ser depurada, seleccionada e interpretada, puesto que los resultados y las conclusiones no siempre resultan evidentes ni exentos de problemas. Cuando se hace ciencia empírica se puede obtener información muy precisa: dataciones fósiles, estructuras cerebrales, localizaciones y registros de neuronas individuales, observaciones conductuales, etcétera, pero sin el debido cuidado en el proceso inferencial, esta información puede dar lugar a teorías carentes de sustento real. Es por eso que resulta primordial el cuidado en el proceso que va del diseño del experimento a la elaboración de enunciados teóricos a partir de la evidencia obtenida. A este objetivo ayuda en buena medida el trabajo filosófico.

En forma histórica, los filósofos han emprendido el trabajo de investigación de la esencia de la mente en forma teóri-

co-especulativa. Aun aquellos que, como Descartes, Leibniz o Kant colaboraron fuertemente al avance científico de sus respectivas épocas, elaboraron teorías de la mente de corte trascendental. De esta manera, quedaba delimitado el espacio lógico de lo que significaría ser una subjetividad para cualquier individuo posible. La descripción de la mente en las especies reales no comenzó a ser importante sino hasta el surgimiento de las ciencias empíricas, particularmente de la biología y la psicología, durante el siglo XIX. Hasta ahora, hemos revisado cómo la filosofía contemporánea de la mente se ha valido de las ciencias del cerebro, de la biología, de la computación y de la conducta en un mutuo trabajo de cooperación para la elaboración de sus teorías. Esta tendencia es cada vez más importante y, a pesar de que hay de por medio un debate meta-teórico sobre la relación entre filosofía y ciencia empírica para el estudio de los procesos cognitivos, son pocos los filósofos que el día de hoy afirmarían tajantemente que la comprensión completa de la mente y su función proviene exclusivamente del trabajo filosófico especulativo.¹

Ya sea por medio del estudio teórico especulativo o por el empírico observacional, el énfasis siempre estuvo puesto en la descripción del funcionamiento normal de la mente. Kant, por ejemplo, estudia la estructura del conocimiento a partir de una conciencia adulta con facultades de percepción

¹ A este tema se le conoce como el debate sobre la *naturalización de la filosofía*. De igual forma que es difícil encontrar una postura trascendentalista radical, tampoco es fácil encontrar el extremo opuesto del espectro: alguien que afirme, desde la filosofía, la suficiencia absoluta de la ciencia empírica para el estudio de la mente. Esta fue la postura original de Quine (1969) que resultó muy influyente, pero que, con el tiempo, se fue moderando.

normales y de razonamiento infalible. Los casos de ilusiones y de engaños sólo eran abordados como vía de refutación de los argumentos escépticos. No obstante, como veremos a lo largo de este capítulo, los sesgos, déficits, trastornos y enfermedades son tanto o más útiles cuando se trata de comprender la forma en que funcionan nuestros estados mentales. Una persona que pierde cierta habilidad, por ejemplo, la de producir lenguaje hablado, es una importante fuente de información: sobre qué otras facultades se relacionan con dicha función y cuáles son independientes, sobre la dificultad con que se puede regenerar la función o si se trata de una pérdida irreparable, sobre la existencia de posibles grados en el daño y, en última instancia, sobre las estructuras cerebrales implicadas en la función perdida. Desde los inicios de la medicina científica, los médicos y cirujanos conocían las correlaciones entre lesiones cerebrales y déficits gracias, entre otras cosas, a los soldados heridos en batalla. Sin embargo, estos casos no resultaron interesantes para su estudio filosófico sino hasta décadas recientes. Hoy en día, esta vía negativa e indirecta de información se ha consolidado en una nueva rama filosófica: la *filosofía de la psicopatología*.

En el presente capítulo, llevaremos a cabo algunas precisiones metodológicas y teóricas sobre lo que puede significar hacer una filosofía de la psicopatología. Posteriormente, se presentarán algunos trastornos mentales que pueden ser de relevancia filosófica. Finalmente, expondremos los debates teóricos y los retos que presenta la comprensión de la estructura de la mente emanada de dichos trastornos para las teorías filosóficas tradicionales.

1. Filosofía de la psicopatología

Hemos afirmado que la *filosofía de la psicopatología* es una nueva disciplina filosófica. Sin embargo, como sucede con toda nueva disciplina, hay múltiples sentidos diferentes que están implícitos detrás del término usado. Podemos señalar con claridad tres de esos sentidos, cada uno de ellos con un objetivo distinto. La filosofía de la psicopatología puede referir:

1. Al objetivo de hacer una filosofía de la ciencia psiquiátrica. En tanto filosofía de la ciencia, esta disciplina se encargaría de estudiar los problemas relacionados con la cientificidad de la psiquiatría: delimitación, explicación, relaciones inter-teóricas, leyes, modelos, tradiciones de investigación, etcétera. Se trataría de un apartado específico de la comprensión filosófica de la ciencia, en el mismo nivel que la filosofía de cualquier otra disciplina científica.
2. A la discusión sobre la pertinencia ética y práctica de la taxonomía psicopatológica y de la práctica psiquiátrica en general. Una disciplina así tendría como tarea evidenciar y criticar los supuestos epistémicos que subyacen a la práctica del psiquiatra. Antecedentes de este trabajo los podemos encontrar en autores como Foucault o Canguilhem. Más recientemente, bajo esta idea se han emprendido proyectos de reconsideración de la enfermedad mental con perspectivas no ortodoxas: la anti-psiquiatría (Szasz, 2010), la psiquiatría crítica, la psiquiatría con perspectiva de género, etcétera.

3. Al proyecto de interacción entre filosofía de la mente y psiquiatría en un doble sentido: usando los casos psicopatológicos para poner a prueba ideas y teorías filosóficas, y ejercitando el análisis filosófico para traer claridad al diagnóstico y clasificación de los trastornos mentales, al tiempo que se usan las distinciones teóricas como guía para la observación y la experimentación.

En realidad, ninguno de estos tres sentidos es del todo excluyente. Los autores que trabajan actualmente en esta disciplina llevan a cabo tareas que se pueden inscribir en cualquiera de los tres objetivos. No es posible llevar a cabo una crítica sobre la medicalización de la enfermedad mental sin antes abordar temas como los de la explicación o la relación entre teoría y práctica dentro de la psiquiatría. De la misma manera, una interacción entre filosofía y psicopatología ha de traer como consecuencia alguna forma de reconsideración de los límites de ambas disciplinas. En este capítulo haremos énfasis en una filosofía de la psicopatología en el tercer sentido: el de la interacción. Para ello llevaremos a cabo algunas especificaciones metodológicas previas.

En primer lugar, hay que poner atención en el objeto de nuestra investigación: la enfermedad mental. Hasta ahora he utilizado en forma un tanto ligera el término de «enfermedad», pero esta forma de caracterizar aquello que le pasa a un paciente psiquiátrico es, cuando menos, debatible. Acostumbramos pensar que todo aquello que afecta a la salud o al bienestar de una persona es una enfermedad. Creemos que es justo eso lo que define que una persona sea un paciente. Sin embargo, en la literatura especializada hay varios términos que se usan para referir a las afecciones mentales (algunos de

los cuales están mejor delimitados que otros): enfermedad, trastorno, síndrome, desorden.²

Según la OMS,³ para que una afección sea considerada una *enfermedad* debe cumplir cuando menos dos de los siguientes cuatro criterios: 1) tener una causa reconocible o supuesta, 2) presentar un conjunto identificable de signos y síntomas,⁴ 3) tener alteraciones anatómico-fisiológicas consistentes y 4) un tratamiento curativo o paliativo existente o posible. Como se puede ver, el énfasis está en la identificación de la causa orgánica de la alteración de la salud. Si tal causa es identificada o identificable, entonces se asume que el malestar es de origen orgánico y, por lo tanto, tratable en principio. En cambio, un *síndrome* es un conjunto de signos y síntomas con causa (etiología) desconocida. Se trata no de una afección única y específica, sino de un cuadro clínico que puede estar vinculado con uno o varios problemas que, a su vez, pueden ser enfermedades. Pero la existencia de una enfermedad no es necesaria: es el caso del síndrome de Estocolmo. Los signos y síntomas en un síndrome no siempre son estables: puede haber un rango amplio en la manifestación del cuadro clínico de un mismo síndrome, además, tales signos y síntomas pueden variar o desaparecer por completo

² Sobre los problemas filosóficos para especificar en qué consisten los desórdenes mentales, se puede consultar el libro de Graham (2010) *The Disordered Mind. An Introduction to Philosophy of mind and Mental Illness*.

³ Véase: <https://www.who.int/es>

⁴ La diferencia entre signo y síntoma es que mientras que el primero es objetivo, el segundo es subjetivo. El signo es reconocible y medible por un observador (el médico): temperatura, enrojecimiento, supuración, vómito, diarrea; el síntoma es la manifestación por parte del paciente de un malestar que sólo él puede reportar: mareo, náusea, dolor de cualquier tipo, comezón, o, para el caso, alucinaciones o delirios.

sin dejar secuelas. En resumen, la diferencia principal entre enfermedad y síndrome es la indeterminación etiológica (causas y manifestación) del segundo sobre la primera.

Los términos de *trastorno* y *desorden* en realidad no tienen un significado distinto del de síndrome. En ambos se pone énfasis en el aspecto anómalo de la condición del sujeto: la persona presenta una diferencia clara, relevante y conflictiva en su comportamiento respecto de su grupo social. Debido a que su etiología orgánica no es clara, estos tres términos se usan principalmente para las afecciones mentales. Más que aludir a explicaciones, son términos que sirven para describir estados globales de anormalidad mental y conductual en la relación del individuo con su entorno y consigo mismo. En cualquier caso, ninguno de estos términos se compromete con ninguna forma de reducción causal al cerebro. Y es por eso que tales afecciones son de relevancia para la filosofía de la mente.

Ahora bien, una vez identificado el tipo de afecciones que nos interesan, ¿cómo se clasifican? Ese es otro problema metodológico. Existen dos clasificaciones de trastornos mentales que se actualizan periódicamente y que los profesionales de la salud usan como referencia para el diagnóstico y el tratamiento: el DSM (*Diagnostic and Statistical Manual of Mental Disorders*) y el CIE (Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud). Mientras que el DSM es un manual específico para enfermedades mentales, el CIE es una clasificación de toda enfermedad en general; el apartado que nos interesa del CIE, el de los trastornos mentales, corresponde al capítulo V. Otra diferencia es que el DSM está elaborado por la Asociación Estadounidense de Psiquiatría,

razón por la cual es de amplio uso en Estado Unidos y el mundo anglosajón, mientras que el CIE está elaborado por la Organización Mundial de la Salud, de modo que es el manual recomendado internacionalmente por la ONU. No obstante, ambos siguen principios de descripción y clasificación similares, dando como resultado taxonomías más o menos consistentes. Como mencioné, ambos son revisados periódicamente. La versión más reciente del DSM es la V^a (2013), mientras que la del CIE es la X^a (1992).⁵

Con tales manuales, ¿está ya resuelta la cuestión de qué sean y cómo se clasifican las enfermedades mentales, de modo que los filósofos podamos acudir sin preocupaciones a esas fuentes para hacer filosofía de la mente? La respuesta es, claramente: no. De inicio, se trata de manuales de ayuda para el profesional de la salud en el diagnóstico y tratamiento; no son diccionarios ni tratados teóricos sobre los conceptos relacionados con los padecimientos mentales. Si uno acude a estos manuales como vía para enterarse de las diferencias conceptuales esperando encontrar una taxonomía clara y discreta, el resultado será francamente decepcionante. Los grupos más generales de las clasificaciones no son necesariamente del mismo rango categorial; los términos relevantes no están definidos, sino caracterizados clínicamente; conceptos que en ocasiones pueden aparecer como síntomas de un trastorno en otros casos pueden aparecer como trastornos ellos mismos; hay trastornos que no son reconocidos como tales en todos los casos; las caracterizaciones clínicas son

⁵ Las ediciones del CIE se revisan cada tres años. No obstante, ya está en proceso de publicación la siguiente edición, CIE-11, que deberá salir en el 2020.

porosas, es decir, es perfectamente posible que un mismo caso pueda ser diagnosticado como uno u otro trastorno por encontrarse con signos y síntomas limítrofes entre dos categorías no excluyentes; dado que la noción de padecimiento mental depende de una perspectiva específica de la medicina, hay debates importantes sobre la validez transcultural de las clasificaciones. Además, hay que añadir la historicidad de las clasificaciones: en cada revisión hay trastornos que se añaden y otros que se eliminan, y esta selectividad no necesariamente depende de valoraciones epistémicas.

Finalmente, hay que señalar los problemas que surgen al intentar obtener datos de los trastornos mentales en seres humanos. En el apartado sobre las neuronas espejo, mencionamos las dificultades para extraer inferencias analógicas de los hallazgos en macacos y en seres humanos. Y es que la experimentación en seres humanos, particularmente en lo concerniente a padecimientos mentales, tiene límites éticos estrictos. En principio, no es posible realizar ninguna observación, tratamiento o experimento que no tenga como objetivo beneficiar de alguna forma al sujeto o del que, cuando menos, se encuentre suficientemente informado sobre las condiciones, procedimiento y consecuencias. Además, hay que tener presente que al estudiar trastornos estamos hablando de pacientes: sujetos con enfermedades en estados físicos y emocionales alterados, en ocasiones gravemente, por sus respectivos padecimientos. Esto obliga al experimentador a mantener presentes principios de beneficencia y empatía en el trato. Además, la información que se pueda obtener por vía verbal en ocasiones se debe tratar con mucha reserva, sobre todo cuando se trata de padecimientos en donde el paciente carece de la conciencia de su condición (*anosognosias*). Una

última consideración es el problema de la base empírica y la reproducibilidad de las observaciones. A pesar de que podemos caracterizar con cierta claridad los tipos de trastornos y sus cuadros clínicos, cada trastorno, cada lesión cerebral y cada paciente presenta un conjunto de condiciones únicas. No es posible encontrar múltiples casos similares del mismo trastorno con el fin de aislar y probar una variable relevante. Nos tenemos entonces que restringir al estudio de casos paradigmáticos. De ese modo, hay que asumir que no es posible derivar generalizaciones empíricas fuertes de la forma en que sí es posible hacerlo en otros campos de la ciencia.

2. Delirios

El espectro de los trastornos mentales es amplio. No obstante, algunos de ellos presentan retos más interesantes para la reflexión filosófica que otros. En particular, resultan relevantes aquellos padecimientos que tienen que ver con nuestra percepción del mundo, del tiempo y del espacio, con la alteración de nuestras creencias o con la alteración de la conciencia, del propio cuerpo y del sentido del yo.

Los trastornos relacionados con los delirios son quizá los que más atención filosófica han recibido. Esto se debe a que el delirio involucra temas clásicos de la epistemología: la percepción normal y alterada, la postulación y justificación de las creencias y su relación con la evidencia, la representación mental que el sujeto tiene del mundo y de sí mismo,

etcétera.⁶ El problema es que, como sucede con la mayoría de los conceptos mentales, en el concepto de delirio confluyen sentidos cotidianos y técnicos que dificultan su tratamiento. En ocasiones se puede llegar a confundir con términos relacionados como «ilusión», «alucinación» o «engaño». En forma técnica, el delirio puede definirse como una creencia falsa para la que no hay evidencia y que, de hecho, se sostiene recalcitrantemente como verdadera a pesar de que la evidencia en contra sea contundente. El delirio en sí mismo no es un trastorno, sino que es síntoma de uno de tipo psicótico. Los *trastornos psicóticos* son aquellos en donde el sujeto presenta una alteración grave de la percepción y de sus procesos racionales de modo que pierde el sentido de realidad. Ejemplo típico de un trastorno psicótico es la esquizofrenia, en la cual acostumbra haber delirios y alucinaciones constantes.⁷

Los delirios pueden ser leves, moderados o graves (extravagantes); pueden ser temporales, episódicos o permanentes; pueden variar en su tipo: persecutorios, de grandeza, erotómanos, somáticos, nihilistas, etcétera. Como parte de un trastorno psicótico, generalmente los delirios van acompañados de otros síntomas, como alucinaciones y pensamiento desordenado, lo cual crea cuadros clínicos de tratamiento complejo.

⁶ Para informarse sobre la filosofía de los delirios, véase: Bortolotti (2010, 2015, 2018)

⁷ Si aceptamos sin mucho debate la definición técnica de delirio que aquí he dado, entonces la diferencia entre delirio y alucinación sería que mientras que el primero es una creencia falsa, la segunda es una percepción anómala generada sin ningún estímulo externo. Sin embargo, como se verá, incluso la definición de delirio como una creencia es debatible.

Retos filosóficos del delirio

En primer lugar, hay que mencionar el reto filosófico de la especificación conceptual del delirio. Caracterizar a este síntoma como una «creencia falsa recalcitrante a la evidencia contraria» parece no bastar para excluir casos no patológicos. El fervor religioso, por ejemplo, cabría sin problemas dentro de la definición de delirio y, sin embargo, no acostumbra clasificarse como una conducta patológica (no siempre, cuando menos). Habría que poder dar con una caracterización que nos permitiera diferenciar entre una mente delirante y la simple necedad en la conservación de una creencia que el sujeto no está dispuesto a abandonar. Una posible solución (a la que recurren las clasificaciones usuales) sería excluir las creencias recalcitrantes obtenidas por influjo social, como las religiosas. No obstante, nada impide que se desarrollen conductas delirantes a partir de creencias sociales o, en forma inversa, que creencias delirantes que surgieron por un trastorno específico en un sujeto se terminen convirtiendo en creencias sociales de un grupo determinado, como puede suceder en un dogma religioso o político.

Otro problema surge con la noción de «creencia falsa». Un paciente puede padecer un delirio de persecución en el que afirme ser vigilado o perseguido y, no obstante, ser en efecto vigilado o perseguido sin que lo sepa (por ejemplo, por la policía). En la epistemología contemporánea sabemos por los contraejemplos de Gettier (1963) que es perfectamente posible este tipo de coincidencias circunstanciales. Tales ejemplos niegan la relación causal entre el hecho en sí (ser vigilado) y la creencia del sujeto. Pero, en ese caso, por excéntrico que parezca, tendríamos una creencia verdadera

en un sujeto al que no quisiéramos dejar de etiquetar como delirante. El reto sería demostrar que, en los casos delirantes, el sujeto obtuvo la creencia por una vía diferente a la constatación del hecho (el cual puede, incluso, existir). Pero la discusión contemporánea nos ha enseñado que eso resulta bastante más complicado de lo que parece. Pedir una conexión causal de ese tipo para todas nuestras creencias con el fin de demostrar que no son delirantes resultaría poco menos que absurdo. De la misma forma, podemos pensar en casos de delirios que tuvieron su origen en hechos reales constatados por el sujeto. Tratar al delirio como una creencia falsa nos puede enfrentar incluso con los mismos problemas paradójicos que tan bien conocemos en la filosofía: ¿qué hacemos con un paciente delirante que afirma recalcitrantemente que está mentalmente enfermo?, ¿le creemos? En ese caso estaría sosteniendo una creencia verdadera, en cuyo caso ya no podríamos llamarlo delirante, pero si no es delirante, entonces en efecto está enfermo ¿de qué padecimiento?, ¿delirio?... y así sucesivamente.⁸

Quizá entonces el delirio no es tanto una creencia falsa, sino una actitud hacia una creencia determinada. La disposición a no abandonar la creencia a pesar de toda evidencia contraria, elaborando explicaciones inverosímiles pero coherentes podría ser una vía de caracterización de este padecimiento. Los pacientes delirantes con frecuencia tienden a mostrar procesos coherentes y racionales de argumentación a favor de sus afirmaciones. El problema radica en la inverosimilitud del cuadro resultante. No es posible justificar en forma verosímil afirmaciones del tipo «hay alguien que escucha mis pensamientos» o

⁸ Véase Fulford, 1989.

«los pensamientos que tengo han sido insertados en mi mente por alguien más», pero, no obstante, siempre es posible crear una justificación coherente. Esto nos enfrenta con el debate epistemológico de la forma de justificación de las creencias: *fundacionismo* contra *coherentismo*. En forma simplificada, podemos decir que mientras que el fundacionista pone el énfasis en la justificación observacional de creencias básicas para sustentar el resto del sistema en inferencias válidas, el coherentista pretende que el valor epistémico más importante está en la coherencia del sistema en general, de modo que la falsedad de una creencia puede no tener consecuencias catastróficas para el mantenimiento de la red de creencias del sujeto.⁹ El delirio podría ser entonces una forma patológica de justificación coherentista. No obstante, hay que pensar que ningún coherentista sostendría una teoría de la justificación de las creencias que sistemáticamente condujera a afirmaciones falsas o no comprobables sobre la realidad por más coherente que resulte el sistema. Entonces es en el aspecto anómalo de esta forma de justificación en el que habría que poner atención.

Otra posibilidad es que el problema del delirio no radique sólo en su condición de creencia, sino en la relación compleja entre la experiencia y la valoración del sujeto sobre ella. Cuando vemos una manzana roja sobre la mesa y hay una manzana roja sobre la mesa, la relación entre el juicio y el hecho es directa y causal. La alteración de cualquiera de los dos elementos de la relación ha sido el motivo de gran parte de los debates en la historia de la epistemología. Las discusiones sobre las ilusiones, engaños y alucinaciones, así como los grandes sistemas metafísicos tales como idealismo y racionalismo

⁹ Al respecto, véase la compilación de García, Eraña y King (2012).

lismo tienen su fuente en esta relación entre juicio y contenido de la experiencia. Cabe entonces preguntar, el delirio, ¿es un problema del juicio o de la experiencia? Si es el primer caso, entonces lo fallido es la forma en que el delirante sostiene la verdad de su creencia falsa, es decir, es un problema de justificación, como se observó previamente. En el segundo caso, el delirante tendría un problema de carácter sensorial-perceptual que le haría tener una experiencia incorrecta, la cual, a su vez, se vería obligado a justificar en forma estrambótica. Si este fuera el caso, entonces deberíamos poder encontrar evidencia neurofisiológica o cognitiva que explicara la anomalía perceptual. Pero no sucede así. Además, en tales anomalías el sujeto es, con frecuencia, consciente del déficit, consciencia que no está presente en el sujeto delirante.

Quizá el problema radique en ver por separado ambos procesos: el de la experiencia como elemento meramente sensorial y el del juicio como valoración cognitiva de la experiencia. Una posible vía de investigación sería ver el delirio como el resultado anómalo de un proceso complejo que involucra elementos sensorio-perceptuales y racionales-valorativos. Se podría explorar la posibilidad de que el delirio sea la valoración anómala de una creencia muy probablemente falsa producto de una experiencia anómala. Este bucle permitiría explicar cómo es que el delirante entra en estados progresivos de retroalimentación de su propio delirio, es decir, que el sujeto utiliza su propia experiencia anómala como evidencia de sus futuras inferencias patológicas.¹⁰

¹⁰ Esta hipótesis se vería apoyada por casos como los de la esquizofrenia en que la conducta delirante se presenta en conjunción de alucinaciones provenientes de varias modalidades sensoriales: visuales, auditivas, táctiles.

Mencionamos antes que las creencias delirantes pueden ser de varios tipos. En una mente no delirante pueden aparecer esporádicamente estas ideas, pero acostumbran sucumbir a la experiencia y desvanecerse con facilidad. Como ya se dijo, el problema del delirio es la persistencia patológica de dichas creencias. No obstante, algunas de estas ideas han sido el motor original de grandes temas en la filosofía: ¿Y si estoy soñando?, ¿y si soy sólo una mente desprovista de cuerpo?, ¿y si vivo en una simulación computacional programada por un genio maligno?, ¿Y si los demás son impostores (zombies, androides...)? ¿Y si soy yo el impostor que *deambula enmascarado por el teatro del mundo*?, ¿y si en realidad no soy yo, sino alguien más (Dios, por ejemplo) quien controla mis movimientos corporales de modo que no soy responsable de ellos?, ¿y si en realidad mis pensamientos no son del todo privados y es posible que alguien pueda acceder a ellos en alguna forma que desconozco aún?, ¿y si estoy muerto? Para cada una de estas ideas, que para un psiquiatra pueden ser calificadas como delirantes, ha habido cuando menos algún filósofo que las ha considerado temas serios de investigación. Los trastornos que revisaremos a continuación bajo la idea del reconocimiento (del mundo o del propio yo) están basados en delirios que llevan al exceso las ideas filosóficas expresadas en estas preguntas.

3. Reconocimiento del mundo

Cuando caminamos por la misma calle transitada que recorreremos a diario para llegar al trabajo o la escuela somos bombardeados por una gran cantidad de estímulos visuales

y sonoros, principalmente. Observamos edificios, animales (perros, gatos), autos y personas sin prestarles mayor importancia. Nuestros sistemas cognitivos son capaces de recibir la información, valorarla y clasificarla sin mayor problema. Todo eso en un silencioso conjunto de pasos que suceden bajo la superficie de la atención consciente. Pero, en ocasiones, de entre toda esa masa informe y generalizada de rostros hay alguno que resalta entre la multitud: quizá algún amigo, colega o familiar. Hay un instante, breve o largo según sea nuestra familiaridad con dicha persona, en que llevamos a cabo un proceso de reconocimiento. La mayoría de las veces ese proceso es casi instantáneo. Como especie, hemos evolucionado para ser muy eficaces en el reconocimiento de rostros. El proceso nos resulta tan natural que no nos paramos a pensar en la complejidad que eso implica.

Como toda habilidad cognitiva, el reconocimiento de rostros no está exento de anomalías. Imagina que eres capaz de reconocer el rostro de tus familiares o amigos cercanos, pero que «de alguna forma» te resultan ajenos. No habría ningún rastro en particular en esos rostros que te hiciera dudar a quién le pertenecen, pero «hay algo» que te hace pensar que en realidad no es la persona en cuestión... ¡Aunque su parecido sea perfecto! Hacia finales de la década de 1910 un psiquiatra francés identificó este padecimiento y desde entonces lleva su nombre: trastorno de Capgras.

Hoy sabemos que la capacidad de reconocer rostros depende de un sistema neural complejo que involucra áreas de la corteza visual pero también de las zonas asociadas con el procesamiento de las emociones. Cuando reconocemos un rostro no sólo identificamos los rasgos faciales como pertenecientes a una persona determinada, sino que

esa información después es procesada por los centros de la emoción. Ese segundo procesamiento es el encargado del sentimiento de familiaridad que va implícito cuando nos encontramos con un rostro conocido y que nos permite sentir apego o desagrado por cada persona. Es esa valoración emocional de la que carecen los pacientes de Capgras. El resultado es un trastorno muy curioso... y trágico. El paciente es capaz de reconocer el rostro de su familiar o conocido, pero al no tener valoración emocional alguna, extrae la explicación más «sensata»: ¡Se trata de un impostor! Y ahí es donde comienza el delirio. La impronta del déficit emocional es tan fuerte que, por más que el sujeto se percate racionalmente del absurdo que implica que todas las personas a su alrededor hayan sido substituidas por impostores, le resulta imposible no seguir creyéndolo. Es algo similar a las ilusiones que son impermeables a la razón: aunque entendamos cuál es la causa del engaño en la ilusión que nos presentan, no podemos dejar de sucumbir a ella.¹¹ Como los procesos racionales del paciente no se encuentran necesariamente dañados, entonces es capaz de generar explicaciones para la impostura. Como todo delirio, este también es fuertemente recalcitrante a la evidencia. De hecho, a veces resulta contraproducente confrontar al paciente que sufre de Capgras: sus explicaciones son cada vez más inverosímiles y delirantes.

¹¹ Un ejemplo de ilusión de este tipo es el tablero de Adelson, en donde los escaques del tablero de ajedrez se ven de colores diferentes a pesar de ser del mismo tono (https://es.wikipedia.org/wiki/Tablero_de_ajedrez_de_Adelson). Lo asombroso es que, incluso después de saber que se trata del mismo color, seguimos percibiendo los cuadros como colores distintos. Esto es un ejemplo de ilusiones impermeables a la razón.

Como se puede imaginar, es común que este delirio sea fuente de angustia para el paciente. Al final de cuentas, no es grato pensar que todos a mi alrededor han sido substituidos por impostores. Hay quien ha llegado al extremo de matar a sus familiares esperando encontrar al impostor debajo de la «máscara». Pero este delirio no siempre es fuente de sufrimiento. Hay pacientes que mantienen buena relación con sus «impostores». Algunos, por resignación; otros, porque encuentran satisfactorio tener una pareja que siempre parecen conocer por primera vez. El problema, al ser de carácter visual, se complejiza cuando el paciente se mira en un espejo: ¡Hay un impostor en el espejo! Se puede uno imaginar las maromas delirantes que son necesarias para justificar esta creencia una vez que el médico pregunta «¿Y cómo es posible que ese impostor te haya podido substituir?»

El paciente de Capgras falla en reconocer a los seres queridos. Hay, sin embargo, otro extremo de la patología del reconocimiento: el trastorno de *Fregoli*. Leopoldo Fregoli era un hábil actor italiano que fue famoso por su capacidad para interpretar a múltiples personajes en forma vertiginosa durante la misma obra, cambiando de vestuario, voz y gestos de manera casi instantánea. El paciente con este síndrome cree que detrás de todas las personas que le rodean hay una misma persona que, muy hábilmente, se disfraza para vigilarlo y perseguirlo. Así, mientras que Capgras falla en reconocer a las personas detrás de los rostros, para Fregoli el problema es que hay un sobre-reconocimiento de una misma persona detrás de todo rostro. La valoración emocional que está ausente en el primer trastorno está sobreexcitada en el segundo.

Un último caso nos ayudará a tener un cuadro más completo sobre las patologías del reconocimiento del mundo. Se trata de un padecimiento que no es delirante en sí mismo, pero bien puede acompañar a otros procesos que sí lo son. Se trata de la *prosopagnosia*: la incapacidad de reconocer rostros en absoluto. Para el prosopagnósico ningún rostro tiene rasgos distintivos que puedan servir para identificar a las personas. En forma extrema, el paciente ni siquiera puede reconocer un rostro entre otros objetos en una escena. A diferencia de los pacientes con Capgras y Fregoli, este padecimiento se explica por una deficiencia en el procesamiento visual y no en el emotivo. El estímulo ni siquiera pasa del primer procesamiento; es de pensar que, si lo hiciera, la valoración emocional podría funcionar sin problema.

Oliver Sacks¹² menciona un caso grave de prosopagnosia. Un paciente que llega al consultorio a instancias de su esposa pasa todo el interrogatorio sin problemas. Se trataba de un médico culto, inteligente y con una buena conversación. Sacks, al no ver motivo de preocupación, da por terminada la consulta. Pero entonces algo raro sucede: al despedirse, el paciente intenta infructuosamente ponerse el sombrero...sin darse cuenta que lo que estaba agarrando era la cabeza de su esposa. La esposa entonces reveló que era común que lo encontrarán en la calle intentando conversar con los hidrantes para incendios. En este caso, el aplanamiento visual no se restringía solo a los rostros, sino también a los objetos, borrando la distinción categorial básica entre cosa y persona.

¹² Véase Sacks, 1985. Hay que mencionar que el propio Sacks padecía de este problema.

Retos filosóficos en el reconocimiento del mundo

Todos los trastornos reseñados en esta sección tienen que ver con una anomalía respecto al reconocimiento del mundo externo: de las personas en tanto personas (prosopagnosia extrema), de conocidos detrás de rostros (Capgras) o de una misma persona detrás de rostros conocidos (Fregoli). Incluso, cuando el paciente con Capgras es incapaz de reconocerse a sí mismo, es un elemento externo (el espejo) la fuente del delirio. ¿Qué retos arrojan estos trastornos a la reflexión filosófica?

Podemos plantear brevemente cuatro puntos de discusión sobre las patologías del reconocimiento: 1) la identidad personal, 2) la distinción entre razón y emoción, 3) la fuerza argumentativa de las creencias y 4) la cuestión ontológica del mobiliario del mundo.

Sin duda, la falta de reconocimiento en el espejo plantea discusiones sobre la identidad personal y la elaboración de explicaciones delirantes sobre los impostores que rodean al paciente hace surgir el tema sobre lo que nos constituye como las personas que somos para nosotros mismos y la conciencia narrativa: la forma en que constituimos la propia subjetividad a través de nuestra historia vivida. Respecto de la imagen visual como constituyente de nuestra propia identidad, vale la pena incluir en la discusión los casos de las mutilaciones corporales y de los trasplantes, particularmente, los de rostro. Gran parte de las resistencias para el desarrollo de los protocolos médicos que condujeran a trasplantes de rostro exitosos, se debían a consideraciones de tipo psicológico, ¿será posible aceptar un nuevo rostro como propio? La simple pregunta es una muestra de la fuerte impronta de la

visión en la configuración de nuestra idea del mundo. Si bien es cierto que dependemos de la visión para el reconocimiento, también es cierto que somos capaces de reconocernos a pesar de los profundos cambios que el tiempo produce en nuestros rostros: así sucede al ver una foto nuestra de la infancia. La filosofía tiene aquí mucho sobre lo que reflexionar. El concepto de *yo* ha sido explorado desde múltiples aristas, una de las cuales es la de la creación de una unidad narrativa: el *yo* sería una unidad que da sentido cronológico a las experiencias vividas. Si en realidad esas experiencias no son tales, sino falsas creencias delirantes, ¿qué modificaciones deberíamos esperar del *yo* narrativo del paciente?

Sobre este punto, hay que decir, finalmente que la relación con el propio cuerpo es justamente el tema central de las propuestas corporizadas para la definición y estudio de la mente. Entonces, las teorías de este espectro deberían poder dar cuenta de las anomalías en la forma en que sentimos nuestro cuerpo (propiocepción) y en que nos relacionamos con el mundo a través de él (procesos sensorio-motrices). Sobre este problema de la relación patológica con el propio cuerpo, diremos algo más con motivo del tema de la agencia más adelante. Pero, quizá el problema del reconocimiento, más que de percepción es de emoción. Lo que sucede en el paciente que no se reconoce en el espejo es que carece de resonancia emotiva ante un rostro que debería de resultarle muy familiar. Y esto nos lleva al segundo reto filosófico.

Hay que decir que la relación razón - emoción es una discusión de lo más añeja en filosofía. Tradicionalmente, se había visto a la emoción como antagonista de la razón, aquello que nos impedía ser racionales en forma perfecta. Desde la imagen de los dos caballos alados de Platón hasta

el conflicto de razones entre una y otra en Pascal, se veía a ambas facultades como adversarias. La mayoría de las veces esta presunta disputa se interpretaba en términos morales. Sin embargo, los avances de la neurociencia contemporánea nos han llevado a comprender esta relación desde un ángulo distinto: razón y emoción no son antagonistas, sino que colaboran en la valoración, planificación y toma de decisiones.¹³ El problema en los casos de Capgras y Fregoli es una incorrecta comunicación entre los módulos encargados de ambas funciones. La ausencia de valoración emocional o su exceso llevan por igual a conductas delirantes. Esto demuestra que una teoría contemporánea de la racionalidad no puede proscribir a las emociones como un intruso indeseable, sino que debe ser capaz de explicar justamente esta interacción conductual que hace eficiente nuestra vida social. Así, a la luz de esta nueva evidencia, cabe preguntarnos desde la filosofía, ¿en qué medida nuestros estándares de racionalidad deben modificarse para incluir los componentes valorativos propios de la emoción?, ¿cuáles emociones y en qué medida deben ser consideradas?, la relación razón-emoción ¿sólo debe restringirse a la toma personal de decisiones o debe ampliarse para abarcar los procesos de racionalidad colectiva como en el caso del quehacer científico?

Los casos que hemos mencionado son extremos radicales de negación delirante de la realidad. Cuando se les presiona, los pacientes con delirios excéntricos tienden a elaborar explicaciones cada vez más inverosímiles que pueden llegar, incluso, a desafiar las leyes más elementales de la física. Esto

¹³ Al respecto, consúltese lo dicho al final del capítulo 2 con motivo de la pedagogía social.

abre varios debates epistémicos: ¿qué creencias estamos dispuestos a abandonar ante el mantenimiento de la coherencia del sistema de creencias que poseemos?, ¿hay algunas de esas creencias que deban ser menos susceptibles a su falsación con la finalidad de evitar conductas delirantes como las reseñadas?, ¿cuáles?, ¿qué peso hemos de conceder a la evidencia y cuál a la argumentación para la defensa racional de una creencia?, ¿qué debe contar como evidencia? El delirante sigue elaborando explicaciones fantasiosas en contra de la evidencia a pesar de que racionalmente pueda reconocer que sería más sensato abandonar la creencia en cuestión. Y esto se debe a que, incluso en contra de su propio mejor juicio, tiene un sentimiento íntimo y poderoso de certeza a favor de la creencia delirante. Entonces, ¿qué lugar debemos conceder a la certeza como virtud epistémica?¹⁴

Finalmente, está el tema de la división categorial del mundo que se pone en entredicho en la prosopagnosia radical reseñada por Sacks. En la vida cotidiana, estamos involucrados en el mundo en formas que la teoría de la enacción describe como bucles de retroalimentación sensorio-motrices. Estos procesos resultan subpersonales, es decir, no son accesibles a nuestra atención consciente en forma normal. Ese mundo

¹⁴ Esta relación entre creencia y evidencia que se expone en los delirios resulta de obvio interés filosófico. Pero podemos hacer un apunte meta-filosófico al respecto. Las teorías filosóficas son con frecuencia construcciones coherentes de creencias de índole metafísica. Tales sistemas pueden ser lógicamente consistentes sin necesidad de contradecirse con la evidencia empírica disponible. De hecho, pueden ser igualmente recalcitrantes a la evidencia. Pensemos en el *topos hiper uranos* platónico o en la insistencia de ciertos filósofos idealistas radicales a mantener en pie el sistema a pesar de la evidencia en contra: «¡Peor para el planeta!». ¿No podría ser dicha actitud filosófica una forma de conducta delirante?

en el que estamos insertos está ontológicamente amueblado con sus correspondientes etiquetas categoriales: cosas, seres vivos, personas. El conjunto de cosas que está dotado de sentido son las que forman nuestro mundo y ese mundo, unido al conjunto de posibilidades de acción que tenemos en cada circunstancia, forma lo que los fenomenólogos llaman *horizonte de comprensión*. Ante cada grupo categorial tenemos un *pool* de acciones disponibles (*affordances*) que se pre-seleccionan al percibir un objeto como perteneciente a una determinada categoría. No se trata de un proceso de interpretación de lo que observamos: salvo en ilusiones perceptuales, difícilmente confundimos objetos con animales o personas. Es tan rápida la identificación de categorías ontológicas en la vida cotidiana que podríamos hablar de una percepción directa e inmediata. No es un proceso racional. El problema viene al notar que los límites de dichas clasificaciones son porosos. No siempre está tan clara la etiqueta que corresponde. El sinsentido proviene precisamente del pasmo al no saber cómo clasificar el objeto en la escena que veo. Lakoff y Johnson (1980) ponen como ejemplo la orden: «escala el lápiz» como ejemplo de una frase sin sentido. Que una orden así cause asombro se debe a que no sabemos cómo llevarla a cabo, pues no está dentro de nuestras posibilidades corporales el poder realizarla. Los autores utilizan este y otros ejemplos para demostrar su tesis de la corporización de los significados conceptuales.

Lo que pone de relieve el caso reseñado por Sacks es justamente la posibilidad de que nuestras categorías ontológicas más básicas se difuminen por completo, y que, sin embargo, seamos del todo inconscientes de ello. Las disponibilidades (*affordances*) de interacción física están determinadas en par-

te por la información recibida por los sentidos externos y en parte por el sentido del propio cuerpo (propiocepción). Es comprensible que dicho proceso sea automático e inconsciente, de lo contrario, sería imposible la acción eficaz del sujeto en el medio. Lo que reta la comprensión es la *anosognosia* del caso. Incluso cuando la acción intentada fracasaba (ponerse a su mujer como sombrero), el paciente insistía en llevar a cabo la acción. Esto puede entenderse como una ruptura del bucle sensorio-motriz correspondiente: lo que se ve afectado es la parte del retorno de la acción motriz. La pregunta a reflexionar aquí es: ¿por qué esta ausencia de retorno motriz no tiene como efecto una alerta consciente sobre la imposibilidad de la acción y, con ello, el pasmo que debería surgir ante el resquebrajamiento de la clasificación ontológica con su correspondiente pérdida de sentido?

4. Cuerpo, propiedad y agencia

En un famoso pasaje de *La Náusea* (1938), el protagonista, Antoine Roquentin, reflexiona sobre su propia mano como un ente ajeno, pesado, caliente y vivo en el extremo de su brazo. Aunque reconoce ese apéndice como parte de sí mismo, lo «siente ajeno»; habla de la mano en tercera persona, como teniendo una vida propia y con acciones autónomas. El pasaje de esta novela de Sartre trata de abrir la reflexión sobre la conciencia de la existencia de uno mismo a partir del cuerpo. Aquí, nos sirve para cuestionarnos sobre la relación que tenemos con nuestro propio cuerpo: ¿En qué sentido «es nuestro»? y ¿en qué medida podemos decir que nuestro yo consciente es la causa de sus acciones?

Como producto de derrames y tumores cerebrales, o de callosotomías (operaciones que cortan el cuerpo calloso de un paciente epiléptico), es posible que se presente el curioso *síndrome de la mano ajena*. En este padecimiento, una mano simplemente parece cobrar vida propia, actuando con absoluta independencia de la voluntad del paciente. La mano «toma decisiones» y lleva a cabo acciones en contra de los objetivos de la persona al grado de comprometer su propia vida. Algunos de los casos resultan cómicos: mientras que la persona abotona su camisa con la mano «obediente», la otra hace de las suyas desabotonando insistentemente. Otros casos llegan a lo trágico: la mano rebelde puede pelearse con la otra mano, atacar o, incluso, intentar ahorcar al paciente. ¿Cómo es esto posible?

Otros síndromes y padecimientos muestran consecuencias aún más raras sobre la relación con el propio cuerpo: en algunas versiones del síndrome de Cotard el paciente refiere no tener órganos internos, no tener cuerpo o tener un cuerpo ya muerto;¹⁵ hay casos de daño cerebral en donde el paciente afirma estar moviendo sus extremidades cuando, en realidad, sufre de una parálisis lateral; en otros casos, el paciente afirma que sus extremidades le pertenecen a alguien más: a algún familiar o conocido; en los famosos casos de miembros fantasma, el paciente puede sentir dolor, comezón, o placer en brazos o piernas que han sido amputados;¹⁶ en otros casos hay pérdida absoluta del sentido de la propiocepción, dando como resultado la postración del paciente en cama por no

¹⁵ En otras versiones aún más drásticas, el paciente afirma sentir «como que está muerto» o que, de plano, no existe. Véase, Barrios y Luque, 1995.

¹⁶ Al respecto de los miembros fantasma, se puede consultar el libro de Ramachandran y Blakeslee (1999) *Fantasmas en el cerebro*.

poder moverse sin ver la posición del propio cuerpo. Si bien es cierto que estos casos son producto de diversas causas: derrames, síndromes psicóticos, amputaciones, daño en el área parietal del cerebro, etcétera, todos ellos en conjunto ayudan a delinear las fronteras patológicas de la forma en que el cuerpo le pertenece al *yo*.

Retos filosóficos sobre el cuerpo

¿Cómo está unida el alma al cuerpo?, se preguntaba Agustín de Hipona (De Hipona, 2007). Parece tratarse de una unión íntima a tal grado que parece no haber un proceso de comunicación, sino de conciencia inmediata cuando recibimos una sensación corporal o cuando ejecutamos un movimiento. En ninguno de los dos casos parece mediar un proceso de comunicación como el del capitán que manda a mover las velas del barco o cuando le avisan que el casco de la nave chocó contra un iceberg. Pero, más allá del proceso fisiológico que involucra a diferentes sistemas corporales (nervioso central y periférico y músculo-esquelético), ¿cómo es que se realiza esta unión íntima? Parece ser que dicha inmediatez es una ilusión retrospectiva. Los estímulos sensoriales y las órdenes motrices tardan un tiempo determinado en recorrer los canales nerviosos correspondientes. Para acortar los tiempos de reacción en casos de urgencia, los estímulos toman un atajo que no requiere pasar por el sistema nervioso central: así funciona el arco reflejo cuando tenemos que alejar la mano del fuego, por ejemplo.¹⁷ Pero aún en esos casos, el

¹⁷ Dicho arco reflejo ya había sido descrito por Descartes en su obra *Tratado del hombre* (2014).

proceso no es instantáneo. Esa sensación de inmediatez, por tanto, pareciera una adjudicación *a posteriori*.

Pero la pregunta misma sobre la forma en que se relaciona la conciencia con el cuerpo, establece ya una hipótesis dual: hay dos elementos cuya relación, por estrecha que sea, hay que intentar explicar. Desde la filosofía, podemos preguntar, ¿esta hipótesis subyacente sigue siendo acertada hoy en día dados los conocimientos científicos que poseemos sobre el cuerpo y la mente?, Cuando hablamos de la mencionada relación tendemos a asumirnos del lado de la mente, preguntando cosas como «¿cómo es posible que yo me haga consciente de *mi propio cuerpo*?». Si la relación es evidentemente tan estrecha ¿por qué seguimos asumiendo que el cuerpo es una propiedad y no parte constitutiva del yo?

Esto nos lleva a la pregunta sobre la localización de la conciencia: ¿la conciencia tiene un lugar específico desde el cual controla lo que le sucede al resto de mi cuerpo?, ¿o está distribuida a lo largo de...dónde? Plantear hipótesis sobre la esencia de la conciencia requiere incluso tomar postura sobre el debate de qué cosa sea ésta: ¿un lugar a donde llegan los estímulos y de donde parten las órdenes (tal es la respuesta cognitivista clásica)?, ¿un proceso perteneciente al individuo en su conjunto y no a una parte de él (tal es la respuesta de algunos funcionalismos)? O, ¿un proceso complejo corporal que se extiende al medio por la interacción con las herramientas de las que nos valemos (este es el espectro de las teorías corporizadas, extendidas y enactivas)?

Ahora bien, más allá de cómo entendamos el vínculo entre la conciencia y el cuerpo, podemos preguntar: ¿puedo siquiera pensar en no tener un cuerpo? Los casos patológicos en donde el paciente presenta delirios de negación

de su propia corporalidad nos pasman por la contradicción patente. No obstante, en términos analíticos, aceptamos la argumentación cartesiana para demostrar la inmediatez de la mente sobre el cuerpo y, con ello, su teoría dualista substancial. ¿Qué conclusiones podemos sacar de la reflexión cartesiana que afirma la posibilidad lógica de pensar la mente sin necesidad de un cuerpo y de la afirmación empírica del paciente que afirma ser un *yo* descorporizado? Dado que los casos en que se presentan delirios de descorporización son relativamente comunes en ciertos trastornos, una tarea filosófica relevante es la de desentrañar qué aspectos de la experiencia se encuentran alterados en los pacientes de modo que dichos delirios sean posibles. ¿En qué nivel se encuentra el problema?: ¿en el meramente fisiológico, evidenciando una anomalía de carácter sensorial, o en un nivel de procesamiento superior, ya sea en la elaboración perceptual o en el juicio sobre la propia experiencia?, ¿de qué es exactamente de lo que carece la cualidad subjetiva de la experiencia del sujeto delirante para afirmar que su cuerpo no le pertenece o que, incluso, no posee un cuerpo?¹⁸

Finalmente, tenemos la cuestión de la agencia. La agencia es la facultad de un sujeto para actuar de acuerdo a un objetivo. La filosofía de la acción es una de las ramas de la disciplina más activas en la actualidad. Tiene consecuencias y resonancias en la ética, el derecho, en la teoría de la racionalidad y, por supuesto, en la neurociencia cognitiva.

Nuestro cuerpo puede ser movido, al igual que cualquier otro objeto físico, por una fuerza externa (cuando uno es

¹⁸ Agradezco la ayuda de Andrea Guerrero sobre lo relacionado con el síndrome de Cotard y los delirios de negación del cuerpo.

empujado fuera del vagón del metro, por ejemplo). Pero también puede ser movido en forma endógena: por el deseo voluntario de moverlo con un objetivo; a esto último le llamamos *acción*. En esos casos, hay la intención de lograr un objetivo (salir del vagón del metro) y un movimiento que lo realiza (dar el paso). La conciencia de que ese movimiento de *mi* cuerpo corresponde a esa intención que es *mía*, y de que el primero es causado por la segunda, es la que produce el sentimiento de agencia: soy yo el responsable de la acción. El problema viene cuando, tal como sucede con los casos de la mano ajena, el cuerpo parece moverse con intención, pero sin el consentimiento del sujeto: la mano se mueve no sólo sin la intención previa del sujeto, sino incluso en contra de ella (con una intención propia). En esos casos hay una conducta intencionalmente orientada y otra que se le rebela. La segunda no es reconocida como propia por el paciente. ¿Qué dice eso sobre nuestra teoría de la acción?, ¿debemos entender la segunda acción como desprovista de intención alguna o como motivada por una intención que pasa inadvertida a la conciencia del sujeto?, en ese segundo caso, eso significaría que hay componentes esenciales de la acción que no son accesibles a la conciencia ¿cuáles son?, ¿cómo trabajan? Si a la mano ajena no se le reconoce como propia y, por tanto, tampoco se reconoce que sus movimientos sean consecuencia de la voluntad del sujeto, entonces ¿deberían calificarse sus movimientos como «acciones»? En caso afirmativo, ¿sigue siendo el paciente ética y legalmente responsable de lo que haga esa extremidad? En otras palabras, el caso de las manos que atentan contra la integridad de los propios pacientes ¿deberían ser tratados como intentos de homicidio o de suicidio?

Esto revela que el sentido del yo no es unívoco ni sencillo. Cuando decimos «yo» estamos aludiendo a una gran cantidad de aspectos distintos:¹⁹ agencia y propiedad son dos de ellos, pero también historicidad, unidad perceptual, autoconciencia, propiocepción, la identificación gramatical en primera persona, etcétera. Una tarea del filósofo es, por ejemplo, la de desentrañar estos diferentes aspectos del sentido del yo e intentar explicar sus relaciones dinámicas y lógicas: cuáles son constitutivas para qué funciones, cuáles se relacionan necesariamente entre ellas y cuáles sólo en forma accidental, etcétera. Por ejemplo, la relación entre propiedad y agencia puede ser usada para explicar ciertos delirios: Zahavi y Gallagher (2008), utilizan esta distinción para analizar el delirio de escuchar voces ajenas que es propio de la esquizofrenia. Si bien en ese trastorno se conservaría la propiedad: «la voz está en mi cabeza», lo que estaría ausente sería la agencia: «esa frase no la produce yo». De esta manera podemos usar la labor analítica en la filosofía para derivar ideas que nos ayuden a comprender mejor al mismo tiempo un concepto tan central a la filosofía como el del «yo», así como los casos patológicos aquí referidos.

5. Conciencia

Cerramos este texto con el tema más fascinante de la filosofía de la neurociencia cognitiva: el de la conciencia. Pocos

¹⁹ Respecto de los múltiples sentidos del yo que se pueden distinguir en la filosofía analítica contemporánea, véase Gallagher, 2011.

conceptos son más problemáticos para definir. Mucha tinta ha corrido en la filosofía, y recientemente en la ciencia, sobre la definición, esencia, funcionamiento y estructura de la conciencia. No obstante, todos reconocemos ser seres conscientes en forma íntima e inmediata; ni siquiera es necesario reflexionar sobre ello para saberlo. Es la conciencia misma la que nos permite saber de su propia existencia: se exhibe en su propio acto. Es por eso que, desde que Thomas Nagel la reintrodujo al debate en la filosofía de la mente en 1974, el problema de la conciencia se ha convertido en el santo grial que toda persona interesada en la filosofía de la mente y de la ciencia empírica intenta conquistar. En este breve texto no es posible hacer un tratamiento amplio sobre este tema, pero sí podemos abordar algunos de sus problemas a partir de un conjunto de casos patológicos que nos ofrecen una ventana al funcionamiento y estructura de la conciencia humana.

¿De qué somos conscientes? Si apelamos al concepto cotidiano de conciencia como sinónimo de «percatarse», entonces la pregunta puede parecer de una obviedad pasmosa: es *obvio* que somos conscientes de nuestro entorno, de nuestro cuerpo y, por supuesto, de nuestra propia existencia. Sin embargo, hemos revisado algunas patologías que ponen en entredicho esta presunta obviedad. Capgras, Cotard, el síndrome de la mano ajena, o las hemiplejias nos muestran que nuestra conciencia, cuando menos en dicho sentido cotidiano, es bastante más frágil de lo que queremos suponer. Por lo que hemos visto, la conciencia es una cualidad presente en la mayor parte de nuestros estados mentales cotidianos que, no obstante, puede verse fácilmente alterada en múltiples padecimientos. Pero aún más interesante resulta que nuestra conciencia parece ser una cualidad recursiva:

podemos hacernos conscientes de que somos conscientes. Gracias a esta característica es que podemos identificarnos como sujetos autoconscientes. Pero es de asumir que en este paso reflexivo también tenemos posibilidad de encontrar padecimientos. Cuando un paciente no es consciente de su propio déficit, tenemos una *anosognosia*. Como ha quedado en evidencia en algunos de los casos que aquí he reseñado, las anosognosias no son propiamente un tipo de padecimiento, sino una característica de algunos de ellos.

Los casos de pérdida de memoria acostumbran ser herramientas muy útiles para delinear las fronteras de la conciencia. Los pacientes con daño en el lóbulo temporal tienden a perder la facultad de consolidar recuerdos a largo plazo. Para ellos, la vida está siempre en tiempo presente de forma novedosa y persistente, pues al no poder almacenar la información recibida, todo se presenta por primera vez: personas, sucesos... ¡y hasta el propio rostro en el espejo! Al no hacerse conscientes de su déficit, no se percatan de lo que no son capaces de hacer: recordar. Esto lleva a un deterioro asombroso de uno de los sentidos del yo: su aspecto narrativo que sirve para unificar las experiencias pasadas con los planes futuros en una unidad funcional. Uno de estos pacientes, cada día dejaba constancia en su diario de su «despertar» ante el mundo. Como si antes de eso hubiera estado dormido o aletargado. El problema es que estos despertares se sucedían todos los días sin siquiera recordar todas las recurrencias previas. Sólo al ver que en la entrada anterior del diario refería ya «haber despertado», entonces se veía en la necesidad de escribir «ahora sí, *realmente*, he despertado» o algo semejante. Al no

existir el pasado, lo único que queda es la manifestación del presente con total patencia.²⁰

Los accidentes cerebro-vasculares pueden dejar *heminegligencias* como secuela. Esta afección consiste en que el paciente deja de ser consciente, es decir, deja de atender, percibir e interactuar con la mitad de su escena visual. Dado que el hemisferio derecho se especializa en el procesamiento de la información espacial, y dado que nuestro cuerpo tiene control contra-lateral, lo más común es que la mitad negada sea la izquierda del campo visual. Los informes médicos sobre pacientes con heminegligencia describen casos en donde la persona deja intacta la mitad izquierda del plato, se deja de rasurar la mitad del rostro u omite saludar o interactuar con nadie al lado izquierdo de su campo visual. Al ser cuestionados al respecto, los pacientes niegan la omisión y tienden a confabular una explicación. Es decir, la heminegligencia es, con frecuencia, también una afección anosagnósica.

Un caso que también está relacionado con la percepción visual es el de los pacientes con la paradójica condición de la *visión ciega* (*blind sight*). La visión ciega no se produce cuando hay daño en el ojo, sino en la corteza visual. El órgano de la vista puede estar en óptimas condiciones, dejando pasar los estímulos luminosos (de ahí que se hable de «visión»), pero el problema está en el procesamiento de esos estímulos

²⁰ Dos de los casos más emblemáticos de la pérdida de memoria (y de la neurociencia en general) son los de H.M. (Henry Gustav Molaison) que perdió la memoria a largo plazo luego de ser operado en la década de 1950 para curar una epilepsia grave; y el de Clive Wearing, un exitoso músico y director de orquesta británico que perdió la memoria a largo plazo debido a una fuerte infección cerebral. Es este último paciente el que, hasta la fecha, sigue escribiendo en su diario sobre sus «despertares».

una vez que llegan a la corteza: la «ceguera» radica en el nivel cortical. El daño puede encontrarse en sitios diferentes de la muy compleja red neuronal que procesa la visión en el lóbulo occipital. Las lesiones en cada una de estas partes producen distintas patologías asociadas con la visión. Por ejemplo, una de esas patologías es la que ya reseñamos con el nombre de prosopagnosia. La que aquí nos interesa en particular, es la asociada con la conciencia del estímulo visual.

Hay una zona que procesa y unifica (no sabemos exactamente cuál ni exactamente cómo) los diferentes componentes de la visión para formar la escena unificada de la que somos conscientes. Si esta zona o proceso se daña, entonces la persona reporta «no ver nada». Sin embargo, aún hay muchos estímulos visuales que permanecen activando las zonas no dañadas de la corteza. En esos estímulos hay información sobre toda la escena: objetos, posición, tamaño, movimiento, velocidad, etcétera. El resultado es una patología muy sorprendente: la persona que afirma no ver nada (porque su escena visual no se completa y, por lo tanto, no se hace consciente) es capaz de «adivinar» qué objeto se le presenta, esquivar pelotas que viajan hacia su cabeza o caminar por un pasillo lleno de obstáculos tan sólo «adivinando» hacia dónde moverse. Pero el procesamiento subconsciente no se limita a cualidades espaciales, algunos pacientes con visión ciega son capaces de reconocer, cuando se les obliga a «adivinar», si una fotografía frente a ellos muestra una cara feliz o triste. El porcentaje de aciertos es suficientemente alto como para deberse al azar. Pero si este dato no bastara como evidencia, se ha mostrado por rastreo del movimiento muscular del rostro que, frente a una foto de una persona feliz, el paciente reacciona contrayendo los músculos relacionados

con la sonrisa; otro tanto con las imágenes de rostros tristes o enojados. Esto sugiere que parte de la información visual pasa directo, sin mediación de la conciencia, a los centros de procesamiento de las emociones.²¹

Retos filosóficos sobre las patologías de la conciencia

Entonces, ¿de qué somos conscientes? A la luz de la evidencia científica, parece que de mucho menos de lo que quisiéramos pensar. No sólo gracias a los casos patológicos que hemos mencionado, sino que, gracias a múltiples experimentos en laboratorios de neurociencia cognitiva, hemos descubierto que hay múltiples maneras de engañar y eludir a la conciencia. Hoy sabemos que es posible enmascarar estímulos de manera que sean recibidos por los sistemas de reconocimiento, y que de hecho determinen nuestra conducta, sin necesidad de pasar por nuestra atención consciente. Esos sistemas son capaces de reconocer rostros, identificar objetos, y entender palabras; todo ello con independencia de nuestra atención. Una a una, la ciencia parece complacerse en derribar nuestras intuiciones más firmes sobre la forma en que trabaja nuestra conciencia. Cada tanto, se descubre un mecanismo pre-consciente detrás de alguna facultad mental que asumíamos como propia de la conciencia.²² Así, si la conciencia viene después de los procesos asociados al reconocimiento, la comprensión semántica, la percepción, el

²¹ Sobre los experimentos de la «visión ciega» (*blindsight*), véase: Holt (2003).

²² Respecto a los experimentos para desentrañar los mecanismos subyacentes a la conciencia, se puede consultar el libro *Consciousness and the brain* de Stanislas Dehaene (2014).

enlace entre múltiples modalidades sensoriales y la toma de decisiones, entonces la pregunta filosófica a afrontar es, ¿para qué sirve la conciencia?, ¿tiene algún uso efectivo (evolutivo) o es un simple epifenómeno producto del funcionamiento cerebral? Si nuestra conciencia no es realmente la causa de nuestros actos, sino que estos se producen por procesos sub-personales inaccesibles en primera persona, entonces, ¿qué implicaciones se derivan respecto al libre albedrío?

Thomas Nagel (1974) definió la conciencia como la cualidad subjetiva de la experiencia o, en forma resumida: *qualia*. Es esa característica de la vida mental que nos hace saber «cómo se siente ser un ser humano»: cómo se siente percibir el sonido de una trompeta o el olor del café, por ejemplo. También afirmó que el problema mente-cuerpo es poco interesante sin integrar la conciencia, pero resulta irresoluble al considerarla. En forma natural creemos que la conciencia es aquello que nos hace humanos y, por tanto, la tenemos en amplia estima. Intuitivamente, sabemos lo que se siente poseerla, pero también tenemos una noción de la infabilidad de la conciencia ajena. No obstante, salvo en discusiones filosóficas o en delirios graves, no tendemos a asumir que las otras personas carecen de vida consciente. Pero, a la luz de lo que hoy sabemos científicamente sobre la conciencia, ¿podemos seguir sosteniendo la primacía y la necesidad de la conciencia? Ese aspecto fenoménico-cualitativo de la experiencia que resulta absolutamente infable en tercera persona ¿es prescindible?, si ese es el caso, ¿no se abriría la puerta a la resolución del problema difícil (*hard problem*) de la conciencia? Existen varios debates filosóficos asociados con este tema: el problema de las otras mentes,

la posibilidad de la Inteligencia Artificial o el debate actual sobre la posibilidad de los zombies filosóficos.²³

Pero quizá el debate filosófico más relevante es el de la forma misma de concebir a la conciencia: ¿qué es exactamente a lo que nos referimos cuando hablamos de estados conscientes?, ¿cuál es la característica esencial de la conciencia? La conciencia fenoménica, tal como la definió Nagel, es posiblemente una de las formas más populares de entender la propiedad en cuestión. Sin embargo, es una definición no exenta de debate. Si los qualia son efectivamente subjetivos e inefables, ¿qué sentido tiene intentar hacer un estudio objetivo y riguroso de ellos? Para otros (Dennett, 1988), la idea misma de qualia debe ser rechazada, pues no existe acuerdo entre los filósofos que usan el término: hay tantos sentidos distintos que hacen al concepto totalmente inútil para la investigación.

Finalmente, hay quien piensa que, aun concediendo que los qualia sean una propiedad real de la conciencia no es ni la más importante y ni siquiera una propiedad necesaria. Entonces, ¿qué otras opciones habría para caracterizar rigurosamente a la conciencia? Hay propuestas de todo tipo: hay quienes piensan que la conciencia se puede analizar en términos de atención, acceso y disponibilidad de la información; para otros se trata de una función de procesamiento de la información de alto nivel de integración; hay propuestas

²³ El experimento mental de los zombies filosóficos se puede plantear en forma de pregunta: ¿es posible la existencia de sujetos conductual y orgánicamente idénticos a los seres humanos, pero sin experiencia consciente? Sobre el tema de los problemas fáciles y difícil de la conciencia y de los zombies filosóficos se puede consultar la extensa obra de David Chalmers, la cual se encuentra reseñada en su página: consc.net.

sobre la conciencia como un polo unificador de experiencias y otras más que aluden a un sentido del yo que, a su vez, podría deshacerse en múltiples componentes: propiedad de la experiencia (*mineness*), propiedad de presentación de la experiencia (*formeness*) y propiedad de ser un sujeto (*meness*).²⁴

Terminemos este apartado aludiendo al tema filosófico de la unidad de la conciencia. En la tradición que va de Descartes a Husserl, la conciencia se ha considerado como una propiedad íntegra y unificadora. Mientras que cualquier objeto físico se me presenta en forma parcial debido a que yo mismo, en cuanto cuerpo, me encuentro en una posición espacio-temporal determinada, los estados mentales se me presentan en forma plena en su inmediatez, sin perfiles o perspectivas parciales. Así, si pienso en un recuerdo o en una idea estos se me presentan completos, no en partes. La conciencia tiene la cualidad de «presentificar» sus objetos, al traerlos a la atención del sujeto. La conciencia está ahí por completo, plena, actual en cada momento que se ejerce. Esta sería la característica integral de la conciencia. Pero la conciencia también se ha visto como unificadora. Para Kant, la mente tiene la cualidad de crear unidades sintéticas a partir de multiplicidades: en la percepción a partir de sensaciones, en el entendimiento a partir de fenómenos y en la razón a partir de conceptos. El último paso de esta cadena de síntesis unificadoras sería la conciencia reflexiva del sujeto mismo en cuanto creador de unidades: *la unidad sintética de la apercepción*. Por lo tanto, la conciencia sería única, indivisible y unificadora. ¿Es así?

²⁴ Para una revisión crítica de estas propiedades véase Zahavi (2014).

Parece que no. Que fenoménicamente, desde la perspectiva en primera persona, parezca que la conciencia funciona de determinada manera no es evidencia más que de su forma de aparición, no de su funcionamiento causal. Hoy sabemos que la conciencia es una propiedad gradual y que puede ser afectada en múltiples sentidos. Uno de ellos, quizá el más radical es el de dividirla. Hasta antes de poder manipular quirúrgicamente el cerebro era impensable partir literalmente la experiencia consciente de un individuo.²⁵ Pero eso es precisamente lo que ha sucedido con las operaciones de *callosotomía*, en donde a pacientes con graves casos de epilepsia de lóbulo temporal se les seccionaba el cuerpo calloso, la estructura que divide los dos hemisferios cerebrales. En apariencia, la operación es exitosa: se controla la epilepsia y parece no haber consecuencias negativas. Pero al realizar estudios más detallados se pudo observar que, si uno logra ser lo suficientemente cuidadoso como para hacer llegar información diferente a cada hemisferio, la conducta era distinta, o incluso contradictoria, en la mitad controlada por cada parte del cerebro. Por ejemplo, se puede hacer llegar información al hemisferio derecho del paciente al mostrar una imagen en su cuadrante izquierdo. El hemisferio izquierdo no ha «visto» nada. Como el izquierdo es el encargado del control del lenguaje, el paciente reportará no haber visto nada. Sin embargo, al pedir al paciente que elija con la mano izquierda (controlada por el hemisferio derecho) un objeto entre varios que se le ofrecen, el sujeto tomará el objeto de la imagen que vio. Es como si su hemis-

²⁵ Es curioso que el término «individuo» aluda, justamente, a la imposibilidad de dividir una unidad funcional sin que pierda sus propiedades.

ferio derecho utilizara la parte del cuerpo que controla para evidenciar lo que vio.²⁶

Ya podemos imaginar la cascada de preguntas filosóficas relevantes: ¿qué vio entonces el sujeto?, ¿de qué es consciente?, ¿sólo se es consciente de lo que se puede reportar verbalmente?, y a todo esto ¿dónde queda la unidad de la conciencia que era presuntamente su cualidad esencial?, ¿con qué otros estados mentales, además de la percepción, es posible «dividir» de ese modo a la conciencia?, ¿con los recuerdos, las decisiones, las emociones?, ¿cuál es la experiencia fenoménica de un sujeto en esas circunstancias?, ¿nuestras teorías epistemológicas deberían tomar en cuenta estos casos junto con lo que hoy en día sabemos empíricamente sobre los límites de la atención consciente para explicar la relación entre conocimiento y conciencia?, ¿qué contraejemplos al estilo Gettier pueden elaborarse al considerar la información que aparentemente conozco pero de la que no soy consciente o, en todo caso, de la que hay dos tipos de conciencia? No puede uno dejar de pensar que, si los filósofos de siglos pasados hubieran tenido a la mano estos casos limítrofes, quizá hubieran descrito a la conciencia de forma distinta. Pero esa evidencia está hoy en nuestras manos, por lo que es pensable que los filósofos de nuestra época son los responsables de elaborar esas nuevas teorías.

²⁶ Sobre los experimentos realizados con pacientes con callosotomía, la mejor referencia es la autobiografía intelectual de quien participó directamente en las operaciones y en los estudios referidos y quien es uno de los iniciadores de la neurociencia cognitiva: Michael Gazzaniga (2015).

CONCLUSIONES

En este libro he tratado de presentar algunos aspectos del área de trabajo de la filosofía de la mente en forma introductoria. La idea detrás de los tres capítulos ha sido exponer la necesidad de estudiar la mente en su complejidad psicofísica. Si el lector ha tenido el deseo de superar el nivel introductorio para aventurarse en la búsqueda de información sobre los múltiples temas que aquí se han mencionado, el libro habrá cumplido ya una parte importante de su objetivo.

Otro objetivo que espero haber logrado es el de superar la concepción de que una sola idea, generalmente de carácter reductivo, podrá dar la explicación absoluta de la mente en cuanto a su esencia y su funcionamiento. Ni las teorías del espectro de la TCRM, ni las corporizadas por sí solas, logran ser esa teoría última y total que esperaríamos encontrar. Es por eso que el tercer capítulo presenta una forma «negativa» de abordaje: la filosofía de la psicopatología. La idea de base es que todos los esfuerzos disciplinarios son útiles en la empresa de comprender lo que, hasta ahora, parece ser el sistema más complejo jamás creado en el universo: el sistema mente-cerebro. Para ello será necesario recurrir a los

esfuerzos reflexivos, analíticos y críticos de la filosofía, pero también a la observación empírica en forma de obtención de datos y de elaboración de modelos en sujetos sanos, así como a la observación atenta de los casos patológicos con el fin de comprender los límites del funcionamiento mental. Como dijimos a lo largo del tercer capítulo: la pérdida de las funciones mentales normales deja un vacío que puede ser delineado con precisión para conocer el funcionamiento correcto de la mente.

La cuestión es que este tema, aun comprendiendo la necesidad de un estudio interdisciplinario complejo para entender el sistema mente-cerebro, no es un trabajo ya resuelto y, en muchos casos, ni siquiera hay una ruta clara a seguir. Hay problemas de carácter disciplinario, metodológico, conceptual, epistémico y hasta académico-profesional que hay que resolver. Es ahí donde la profesión filosófica cobra importancia. El trabajo meta-disciplinario requiere salir de las fronteras propias del conocimiento especializado para poder establecer puentes que se traduzcan en métodos y protocolos específicos de colaboración. A su vez, esto requiere del análisis de la historia de las asunciones ontológicas y epistémicas que han dirigido hasta ahora la investigación con el fin de revisar su viabilidad y vigencia con miras a trabajos de mayor amplitud colaborativa.

La crítica de las fronteras de los saberes especializados es tarea del filósofo de la ciencia. Pero el filósofo también ha de colaborar en la consolidación del área atendiendo el rigor conceptual e inferencial que ayude a elaborar teorías sólidas. Finalmente, la filosofía, como vimos en el caso de la filosofía de la psicopatología, puede ella misma colaborar activamente en forma bidireccional con la ciencia empírica proponiendo

interpretaciones y explicaciones de los casos estudiados, al tiempo que puede proponer diseños experimentales basados en la reflexión con el fin de poner a prueba las hipótesis teóricas que le son propias.

El camino es arduo pero prometedor. Pasamos de autores capaces de saber de múltiples temas y escribir sobre todos ellos con fluidez (casos como los de Descartes, Leibniz y Kant son buenos ejemplos en filosofía) a la atomización de los saberes en donde los especialistas se volvieron progresivamente más ajenos entre ellos, para llegar hoy a una época en donde se privilegia el diálogo académico entre diversos profesionales. La cantidad de información disponible hoy en todos los rubros hace indispensable tal colaboración. Por fortuna, también tenemos actualmente los medios necesarios para trabajar en ese sentido. Las páginas, de paga o de libre acceso, con bases de datos de artículos y *journals* de cualquier área son de consulta relativamente fácil. La estructura académica de las universidades e institutos de investigación comienza a moverse para ofertar nuevas carreras que se encuentran entre los límites de profesiones que hasta ahora estaban separadas. De igual modo, se promueve desde todas las trincheras universitarias el estudio de posgrados y la movilidad académica.

Me gustaría terminar este libro con una exhortación a consolidar el papel del filósofo en este nuevo mundo académico. Si bien es cierto que no es prescindible en absoluto el correcto conocimiento de la historia de nuestra disciplina, también es cierto que hoy menos que nunca podemos dejar que nuestra profesión se termine por convertir sólo en historia de la filosofía. Los genios filosóficos que tanto leemos durante nuestra formación fueron, ellos mismos, grandes observadores informados del conocimiento empírico de su época.

ca. Sólo gracias a ese conocimiento y a su capacidad reflexiva es que pudieron elaborar teorías generales que casaban con la evidencia disponible en su época. Hoy esa evidencia es abrumadoramente mayor. El estado de nuestro conocimiento actual demanda una filosofía científicamente enterada para poder comprometerse de lleno con el trabajo de producción y análisis de ese conocimiento. A final de cuentas, la filosofía es reflexión segunda. Si no estamos suficientemente informados del conocimiento que actualmente se produce, corremos el riesgo de hacer de nuestra profesión una disciplina vacía y estéril que, por muy coherente que resulte su trabajo, se encuentre totalmente aislada de los grandes problemas que alguna vez nos interesaron históricamente.

REFERENCIAS

- AIZAWA, K. (2014). The Enactivist Revolution. *AVANT. The Journal of the Philosophical-Interdisciplinary Vanguard*, V(2), 19-42. <https://doi.org/10.26913/50202014.0109.0002>
- AYER, A. J. (comp.) (1965). *El positivismo lógico*. Fondo de Cultura Económica.
- BARTON, G. y Garvis, S. (eds.) (2019). *Compassion and Empathy in Educational Contexts*. Palgrave Mcmillan.
- BARTRA, R. (2007). *Antropología del cerebro: La conciencia y los sistemas simbólicos*. Fondo de Cultura Económica.
- BARNES, B. (1974). *Scientific Knowledge and Sociological Theory*. Routledge & Kegan Paul.
- BENNETT, M. y Hacker, P. (2003). *Philosophical Foundations of Neuroscience*. Blackwell.
- BERRIOS, GE. y Luque, R. (1995). Cotard's syndrome: analysis of 100 cases. *Acta Psychiatr Scand*. 91: 185-188.
- BLOOM, P., y German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of Mind. *Cognition*, 77(1), 25-31.
- BLOOR, D. (1976). *Knowledge and Social Imagery*. The University of Chicago Press

- BORTOLOTTI, L. (2010). *Delusions and other Irrational Beliefs*. Oxford University Press.
- BORTOLOTTI, L. (2015). *Irrationality*. Polity Press.
- BORTOLOTTI, L. (2018). *Delusions in Context*. Palgrave Macmillan.
- BOTTINEAU, D. (2010). Language and enaction. En J. Stewart, O. Gapenne, y E. Di Paolo (Eds.), *Enaction: Toward a new paradigm for cognitive science*. MIT Press.
- BRENTANO, F. (2020). *Psicología desde el punto de vista empírico* (S. Sánchez Magallón, Trad.). Sígueme Ediciones.
- BROOKS, R. A. (1990). Elephants don't play chess. *Robotics and Autonomous Systems*, 6(1-2), 3-15. [https://doi.org/10.1016/S0921-8890\(05\)80025-9](https://doi.org/10.1016/S0921-8890(05)80025-9).
- BROOKS, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47(1-3), 139-159. [https://doi.org/10.1016/0004-3702\(91\)90053-M](https://doi.org/10.1016/0004-3702(91)90053-M)
- BURGE, T. (1979). *Individualismo y mental*. Estudios del Medio Oeste en Filosofía. 4: 73-121
- CARNAP, R. (1932). Psychology in Physical Language. En A.J. Ayer (ed.), *Logical Positivism*. The Free Press, 165-198.
- CHOMSKY, N. (1959). A review of Verbal behavior, by B. F. Skinner. *Language*, 35. 26-58.
- CHURCHLAND, P. S. (1986). *Neurophilosophy: Toward A Unified Science of the Mind-Brain*. MIT Press.
- CHURCHLAND, P. (1989). *Neurophilosophy: Toward a Unified Science of the Mind-brain*, MIT Press
- CHURCHLAND, Patricia. (2016). *El cerebro moral*. Paidós Ibérica
- CHURCHLAND, Patricia y T. Sejnowski. (2017). *The Computational Brain*. MIT Press.
- CHURCHLAND, Paul. (1999). *Materia y conciencia. Introducción contemporánea a la filosofía de la mente*. Gedisa.
- CLARK, A. (1989). *Microcognition*. MIT Press.

- CLARK, A. (2004). *Natural-born cyborgs: minds, technologies and the future of human intelligence*. Oxford University Press.
- CLARK, A. (2016). *Surfing uncertainty: prediction, action and the embodied mind*. Oxford University Press.
- CLARK, A., y Chalmers, D. (1998). The extended mind. *Analysis*. 58. (1), 7-19
- COLOMBETTI, G. (2010). *Enaction sense-making and emotion*. <https://doi.org/10.7551/mitpress/9780262014601.003.0006>
- COSMELLI, D. y E. Thompson. (2010) Embodiment or Envatment?: Reflection on the Bodily Basis of Consciousness. En *Enaction. Toward a new paradigm for cognitivesciences*, MIT Press.
- DAMASIO, A. (1999). *El error de Descartes: La razón de las emociones*. Andrés Bello.
- DAMASIO, A. (2016). *En busca de Spinoza. Neurobiología de la emoción y los sentimientos*. Booket.
- DAMASIO, A. (2019). *La sensación de lo que ocurre. Cuerpo y emoción en la construcción de la conciencia*. (Francisco Páez de la Cadena Tortosa, Trad.). Planeta.
- DAVIDSON, D. (1995). *Ensayos sobre acciones y sucesos*. UNAM.
- DE HIPONA, A. (2007). *Ciudad de Dios. Libros I-VII*. Gredos.
- DE LAPLACE, P. S. (1820). *Ensayo filosófico sobre las probabilidades* (P. Castillo, Trad.). Altaya.
- DEHAENE, S. (2014). *Consciousness and the Brain*. Penguin Books.
- DEMETRIOU, H (2018). *Empathy, Emotion and Education*. Palgrave Macmillan.
- DENNETT, D. (1991). *Consciousness Explained: Little, Brown and Company*. [Trad. al español de Sergio Balari Ravera: *La conciencia explicada. Una teoría interdisciplinar*, Paidós, 1995.]
- DENNETT, D. (2005). *Sweet Dreams. Philosophical Obstacles to a Science of Consciousness*, [Trad. al español de Julieta Barba y

- Silvia Jawerbaum: *Dulces sueños, obstáculos filosóficos para una ciencia de la conciencia*, Katz, Discusiones, 2006]
- DENNETT, D. Marcel, A. y E. Bisiach. (1988). Quining Qualia. En D. Dennett, *Consciousness in Contemporary Science*. OUP, 42-77. [Trad. al español de Ana Isabel Stellino: Quinear los qualia en *La naturaleza de la experiencia. Volumen I. Sensaciones*, Maite Ecurrida y Olbeth Hansberg (comps.), UNAM-Instituto de Investigaciones Filosóficas, 2003]
- DESCARTES, R. (2014). *Meditaciones metafísicas*. Gredos.
- DESCARTES, R. (2014). *Tratado del hombre*. Gredos.
- DI PAOLO y Thompson, (2010). *Enaction: Toward a New Paradigm for Cognitive Science*. MIT.
- DI PAOLO, E., Rohde, M., y Jaegher, H. (2010). Horizons for the Enactive Mind: Values, Social Interaction, and Play. *Enaction: Towards a New Paradigm for Cognitive Science*. <https://doi.org/10.7551/mitpress/9780262014601.003.0003>
- DI PAOLO, E., y Thompson, E. (2014). The enactive approach. En *The Routledge handbook of embodied cognition* (68-78). Routledge/Taylor & Francis Group.
- DI PELLEGRINO G., Fadiga L., Fogassi L., Gallese V. y Rizzolatti G. (1992). *Understanding motor events: a neurophysiological study*. Università di Parma.
- DREYFUS, H. (1972), *What Computers can't do*, Harper and Row.
- EZCURDIA, M., y O. Hansberg (Comps.) (2003). *La naturaleza de la experiencia. Vol. 1: Sensaciones*. UNAM- Instituto de Investigaciones Filosóficas.
- FEIGL, H. (1967). *The Mental and the Physical: The Essay and a Postscript*. University of Minnesota Press.
- FOUCAULT M. (1961). (2006) *History of Madness*. Routledge.
- FOUCAULT M. (1966). *Las palabras y las cosas: Una arqueología de las ciencias humanas*. (Trad. C. Frost). Siglo XXI.

- FOUCAULT M. (1969). *La arqueología del saber*. (Trad. A. Garzón del Camino). Siglo XXI
- FODOR, Jerry. (1975). *El lenguaje del pensamiento*. Prensa de la Universidad de Harvard.
- FULFORD, K. (1989). *Teoría moral y práctica médica*. Cambridge University Press.
- GALLAGHER, S. (2005). Neurophilosophy and Neurophenomenology. En L. Embree y T. Nenon (eds.) *Phenomenology*, 5, Zeta Press, 293-316.
- GALLAGHER, S. (2005). *How the body shapes the mind*. Oxford University Press.
- GALLAGHER, S. (ed.) (2011). *The Oxford Handbook of the Self*. Oxford University Press.
- GALLAGHER, S. y D. Zahavi (2008). *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science*, Routledge.
- GALLESE, V., y Goldman, A. I. (1998). Mirror Neurons and the Simulation Theory of Mind-Reading. *Trends in Cognitive Sciences*, 2(12), 493-501. [https://doi.org/10.1016/s1364-6613\(98\)01262-5](https://doi.org/10.1016/s1364-6613(98)01262-5)
- GARCÍA-CARPINTERO, M. (2007). Funcionalismo. En Fernando Broncano (ed.), *La mente humana*, Trotta.
- GARCÍA, C. L., A. Eraña, y P. King (eds.). (2012). *Teorías contemporáneas de la justificación epistémica. Volumen I: Teorías de la justificación en la epistemología analítica*. UNAM-Instituto de Investigaciones Filosóficas.
- GARCÍA-OCHOA, G y S. McDonald (2021). *Cultural Literacy and Empathy in Education Practice*. Palgrave McMillan
- GAZZANIGA, M. (2015). *Relatos desde los dos lados del cerebro. Una vida dedicada a la neurociencia*. Paidós.

- GETTIER, Edmund. (1963). Is Justified True Belief Knowledge? *Analysis*, 121-123.
- GIBSON, J. J. (1966). *Los sentidos considerados como sistema*. Houghton Mifflin.
- GIBSON, J. J. (1979). *El enfoque ecológico de la percepción*. Houghton Mifflin.
- GOPNIK, A. y A. N. Meltzoff. (1997). *Words, Thoughts and Theories*. MIT Press.
- GOPNIK, A. y H. Wellman. (1994). The Theory Theory. En L. Hirschfield y S. Gelman (eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture*, Cambridge University Press, 257/293.
- GORDON, R. M. (1986) Folk Psychology as Simulation, *Mind and Language*, 1(2): 158–171.
- GOLDMAN, A. I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*, Oxford University Press.
- GOLDMAN, A. I. y L. C. Jordan. (2013). Mindreading by Simulation: The Roles of Imagination and Mirroring, en Simon Baron-Cohen, Michael Lombardo, y Helen Tager-Flusberg (eds.), *Understanding Other Minds: Perspectives From Developmental Social Neuroscience*, Oxford University Press, 448–466.
- GOLDMAN, A. I. (1987). Foundations of Social Epistemics. *Synthese*, 73(1), 109-144. <https://doi.org/10.1007/bf00485444>
- GRAHAM, G. (2010). *The Disordered Mind. An Introduction of Philosophy of Mind and Mental Illnes*. Routledge.
- GRAY, J. (2005). *El reto de Hilbert. Los 23 problemas que desafiaron a la matemática*. (Javier García Sáenz trad.). Crítica.
- HEMPEL, C. (1949). The Logical Analysis of Psychology, en H. Feigl y W. Sellars (eds.), *Readings in Philosophical Analysis*. Appleton-Century-Crofts, 373–384

- HICKOK, G. (2014). *The Myth of Mirror Neurons: The real Neuroscience of Communication and Cognition*. W.W. Norton and Co.
- HOLT, J. (2003). *Blindsight and the Nature of Consciousness*. Broadview Press.
- HUTTO, D. D. (2008). *Folk psychological narratives: The sociocultural basis of understanding reasons*. MIT Press.
- JOHNSON, M. y G. Lakoff. (1980). *Metáforas de la vida cotidiana*. Universidad de Chicago / Ediciones Cátedra.
- JOHNSON, Laird. (2004). La historia de los modelos mentales. En *Psicología del razonamiento: Perspectivas teóricas e históricas*. Editado por Ken Manktelow, Man Cheung Chung. Prensa de psicología.
- KAHNEMAN, D. (2015). *Pensar rápido, pensar lento*. Debolsillo.
- KIM, J. (1996). *Philosophy of Mind*. Westview Press.
- KITCHER, P. (1993). *The Advancement of Science*. Oxford University Press.
- KRIEGEL, U. (2006). Consciousness, Theories of. *Philosophy Compass*. 1(1), 58- 64.
- KUHN, T. (1962). (2013) *La estructura de las revoluciones científicas*. Introducción de Ian Hacking. Traducción e introducción de Carlos Solís. FCE.
- LATOUR y Woolgar (1979). *Laboratory life: The Construction of Scientific Facts*. Princeton University Press.
- LE VAN QUYEN, M. (2010). Neurodynamics and Phenomenology in Mutual Enlightenment: The Example of the Epileptic Aura. En *Enaction*. The MIT Press. <https://doi.org/10.7551/mitpress/9780262014601.003.0010>
- LEVINE, J. (2006). Conscious Awareness and (Self-)Representation. En U. Kriegel y K. Williford (eds.) *Self-Representational Approaches to Consciousness*, MIT Press.

- LINGNAU A., Gesierich B. y Caramazza A. (2009). *Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans*. Proc. Natl Acad. Sci. 106, 9925–9930.
- MAIBOM, H. (2017). *The Routledge Handbook of Philosophy of Empathy*. Routledge.
- MARR, D. y T. Poggio, (1976). Cooperative Computation of Stereo Disparity. *Science, New Series*, 194(4262), 283-287.
- MATURANA H. R. y Varela F. J. (1994). *De máquinas y seres vivos. Autopoiesis: la organización de lo vivo*. Editorial Universitaria / Grupo Editorial Lumen.
- MICHAELS, C. F., y Palatinus, Z. (2014). *Diez mandamientos para la psicología ecológica*. En L. Shapiro (Ed.), *La Routledge handbook de Embodied cognition*, Routledge, 19-28.
- MCDERMOTT, R. y O'Dell, C. (2001) Superando las barreras culturales para compartir conocimientos. *Revista de Gestión del Conocimiento*, 5, 76.
- MOULINES, U. (2011). *El desarrollo moderno de la filosofía de la ciencia (1890-2000)* (X. de Donato, Trad.). Instituto de Investigaciones Filosóficas - UNAM.
- NAGEL, T. (1974). What is it Like to be a Bat? *Philosophical Review*, 83, 435-450. [Trad: al español de Héctor Islas: ¿Cómo es ser un murciélago?, en *La naturaleza de la experiencia. Volumen I. Sensaciones*, Maite Ecurdia y Olbeth Hansberg (comps.), IIF's-UNAM, 2003].
- NAM, C, Nijholt A. y Lotte, F. (eds.), (2018). *Brain-Computer Interfaces Handbook: Technological and Theoretical Advances*. CRC Press.
- NEWELL, A y Simon, H., (1976). Computer sciences as empirical inquiry: Symbols and Search. En *Association for Computing Machinery*, Inc. 19 (3).

- OMS, (1992). *Clasificación Internacional de enfermedades* (10^a ed.). Capítulo V F00_F99: Trastornos mentales y del comportamiento. <http://ais.paho.org/classifications/Chapters/pdf/Volume1.pdf> y <https://www.who.int/classifications/icd/en/#>.
- ONISHI, K. H., Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308 (5719), 255–58.
- PENROSE, R. (2002). *La mente nueva del emperador. En torno a la cibernética, la mente y las leyes de la física*. (José Javier García Sáenz trad.), FCE.
- PINKER, S. (2005). So How Does the Mind Work? *Mind and Language*. 20 (1), 1-24
- PLACE, U.T. (1956). Is Consciousness a Brain Process? *British Journal of Psychology*. vol. XLVII.
- PUTNAM, H. (1967). *Psychological Predicates*. En Capitan y Merrill (eds.) *Art, Mind and Religion*. University of Pittsburgh Press.
- PUTNAM, H. (1975). *The meaning of 'meaning'*. Language, Mind, and Knowledge, K. Gunderson. University of Minnesota Press.
- PUTNAM, H. (1981). *La vida mental de algunas máquinas*. (Martha Gorostiza trad.). UNAM- Instituto de Investigaciones Filosóficas.
- PUTNAM, H. (1983). *Cerebro y conducta*. Rosario Amieva (trad.). UNAM- Instituto de Investigaciones Filosóficas.
- PYLYSHYN, Z. (1984). *Computación y Cognición*. Bradford Books, MIT Press, xxiii, 292
- QUINE, W. (1969). *Epistemology Naturalized*. En *Ontological relativity and other essays*. Columbia University Press.

- RAMACHANDRAN, V.S. y S. Blakeslee. (1999). *Fantasmas en el cerebro. Los misterios de la mente al descubierto*. (Juan Manuel Ibeas trad.). Debate.
- RIZZOLATTI, G. y Sinigaglia, C. (2006). *Las neuronas espejo: Los mecanismos de la empatía emocional*. Ediciones Paidós Ibérica.
- ROWLANDS, M. (2010). *The new science of the mind: from extended mind to embodied phenomenology*. MIT Press.
- RYLE, G. (2005). *El concepto de lo mental*. Eduardo Rabossi (trad.). Paidós, Surcos 4.
- SACKS, O. (1985). *El hombre que confundió a su esposa con un sombrero, y otros cuentos clínicos*. Anagrama.
- SARTRE, J. P. (1938). *La Nausée*. Éditions Gallimard.
- SCHELER, M. (2007). *De lo eterno en el hombre*. Encuentro.
- SEARLE, J. R. (1980). Minds, Brains and Programs en *The Behavioral and Brain sciences*. 3. 417-424. Cambridge University Press.
- SEARLE, J. R. (2014). *Creando el mundo social: La estructura de la civilización humana*. Paidós/ Espasa Libros
- SHAPIRO, L. (2007). *The embodied cognition research programme*. *Philosophy compass*, 2(2), 338-346.
- SHAPIRO, L. (2011). *Embodied Cognition*. Routledge.
- SHAPIRO, L. (2014). *Routledge Handbook of Embodied Cognition*. Routledge.
- SHOEMAKER, S. (1981). Some Varieties of Functionalism. *Philosophical Topics*, 12(1), 93-119. <https://doi.org/10.5840/philtopics198112145>
- SKINNER, B. F. (1957/1981). *Conducta verbal*. Trillas.
- SMART, J.J.C. (1959). Sensations and Brain Processes, *Philosophical Review*. vol. LXVIII.

- SMITH, E. y Semin, G. (2007). Situated social cognition. En *Current Directions in Psychological Science*, 16 (3), 132-135.
- SOLIMAN, T., Gibson, A., y Glenberg, A. M. (2013). Sensory motor mechanisms unify psychology: The embodiment of culture. En *Frontiers in Psychology*.
- STADLER, F. (2013). *El círculo de Viena. Empirismo lógico, ciencia, cultura y política*, Fondo de Cultura Económica.
- STEWART, J., Gapenne, O., y Di Paolo, E. (2010). *Enaction: A New Paradigm for Cognitive Science*. <https://doi.org/10.7551/mitpress/9780262014601.001.0001>
- SZASZ, T. (2010). *The Myth of Mental Illness: Foundations of a Theory of Personal Conduct*. Harper.
- TAPPENBERG, T. (2010). *Foundations of Computational Neuroscience*. Oxford University Press.
- THAGARD, P. (1987). *Computational Philosophy of Science*, The MIT Press.
- THAGARD, P. (2010). *Mind. Introduction to Cognitive Science*, MIT Press. [Trad. *La mente. Introducción a las ciencias cognitivas*, Trad. Silvia Jawerbaum y Julieta Barba, Katz.]
- TURING, Alan, M., (2004). On computable numbers, with an application to the Entscheidungsproblem, Jack Copeland (ed.), *The essential Turing: Seminal writings in computing, logic, philosophy, artificial intelligence, and artificial life: Plus the secrets of enigma*, Oxford University Press.
- VARELA, F., Thompson, E. y Rosch, E. (1991). *The Embodied Mind*, MIT Press.
- VARELA, F. (1996). Neurophenomenology: A Methodological Remedy for the Hard Problem. *Journal of Consciousness Studies*, J. Shear (Ed.), June.
- WATSON, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20, 158-177.

WITTGENSTEIN, L. (2014). *Investigaciones filosóficas*. Gredos.

WOOTTON, D. (2017). *La invención de la ciencia: Una nueva historia de la Revolución Científica*. Editorial Crítica.

ZAHAVI, D. (2014). *Self and Other. Exploring Subjectivity, Empathy and Shame*. Oxford University Press.

ÍNDICE TERMINOLÓGICO

- Acciones epistémicas: 89, 100
- Agencia: 126, 130.
- Anosognosia: 124, 141, 149.
- Armonía preestablecida: 2.
- Arreglo óptico ambiental: 81.
- Autismo: 97.
- autómata probabilístico: 46, 47, 48, 51.
- Brecha explicativa: 85.
- Bucles de retroalimentación: 83, 84, 86, 88, 111, 139.
- Cerebro: 14, 15, 16, 17, 24, 26, 28, 33, 35, 36, 41, 46, 52, 53, 54, 69, 70, 80, 82, 85, 86, 90, 94, 97, 100, 122, 142, 143, 149, 156, 160.
- Cognición social: 88, 93, 102, 108, 109, 112.
- Computación: 23, 46,

ÍNDICE ONOMÁSTICO

- Alan Turing: 46, 57.
- Alfred North Whitehead: 58
- Allan Newell: 57
- Andy Clark: 74, 75, 89, 90, 91, 92, 93.
- Bertrand Russell: 58, 65.
- Brooks: 74, 81.
- Carl Marx: 28.
- Círculo de Viena: 20, 29.
- Cosmelli: 85.
- Daniel Dennett: 73, 154.
- David Chalmers: 74, 75, 89, 90, 91, 93, 154.
- De La Mettrie: 27.
- Di Pellegrino: 94, 98.
- Edmund Gettier: 127, 157.
- Edmund Husserl: 64, 66, 73, 82, 155.
- Engels: 28.
- Euclides: 19.
- Francis Bacon: 27

- 55, 58, 60, 68, 69, 72, 87, 119.
- Computacionalismo: 59, 60, 77.
- Conducta: 37, 38, 40, 41, 42, 43, 44, 45, 46, 49, 54, 56, 58, 89, 90, 101, 102, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 117, 127, 130, 138, 139, 146, 152, 156.
- Conductismo: 18, 23, 26, 37, 38, 39, 40, 41, 42, 43, 44, 45, 49, 52, 53, 55, 89.
- Corporización: 18, 55, 69, 75, 76, 77, 82, 108, 111, 112, 140.
- Cotard: 10, 142, 145, 147, 148.
- CSC (cognición social corporizada): 108, 109, 110, 112.
- Delirio: 125, 126, 127, 128, 132, 133, 134, 136, 138, 149.
- Doble aspecto: 22.
- DSM: 122.
- Dualismo: 22, 26, 27, 77.
- dualista: 21, 26, 52.
- Emociones básicas: 97.
- Enacción: 82, 83, 84, 85, 86, 88, 89, 90, 95, 139.
- Enactivismo: 83, 84, 86.
- Enactivistas: 83, 86, 87, 88.
- Francisco Varela: 73, 74, 82, 83, 85, 86.
- Franz Brentano: 64, 66, 67.
- Galileo Galilei: 27.
- Gallagher: 75, 108, 111, 147.
- George Berkeley: 26.
- George W. Bush: 14
- Georges Canguilhem: 119.
- Giacomo Rizzolatti: 98.
- Gibson: 77, 78, 80, 81, 95, 111.
- Gottfried Leibniz: 22, 36, 117, 161.
- Herbert Simon: 57.
- Hilary Putnam: 57.
- Hubert Dreyfus: 73.
- Humberto Maturana: 82, 85.
- Immanuel Kant: 63, 117, 155, 161.
- Isaac Newton: 19, 27.
- Ivan Pavlov: 40.
- Jean Paul Sartre: 73, 141.
- Jerry Fodor: 57.
- John Haugeland: 58.
- John Locke: 27, 65.
- John Searle: 60, 61, 62, 66, 104.
- John Watson: 40.
- Johnson: 64, 140.

- Enfermedad mental: 119, 120.
- Epistemología: 102, 103, 125, 127, 129.
- Esquizofrenia: 97, 126, 130, 147.
- Estados mentales: 23, 30, 32, 33, 34, 35, 36, 37, 38, 44, 46, 47, 48, 49, 50, 51, 55, 62, 64, 66, 67, 70, 71, 74, 76, 82, 85, 89, 90, 104, 105, 106, 107, 108, 118, 148, 155, 157.
- Estímulos: 40, 44, 94, 95, 97, 114, 131, 143, 144, 150, 151, 152.
- Etiología: 121.
- fenómeno: 19, 29, 37, 72, 84, 88, 155.
- fenomenología: 73, 82.
- Fisicismo: 18, 23, 26, 27, 28, 29, 30, 31, 32, 33, 34, 49, 50, 53.
- Funcionalismo: 18, 23, 33, 45, 46, 48, 49, 50, 51, 52, 53, 59, 84.
- Herramientas fenomenológicas: 65.
- Intencionalidad: 15, 64, 66, 104.
- Materialismo: 27, 28, 29, 33, 34, 36, 37.
- Mente-cerebro: 14, 17, 24, 33, 41, 52, 159, 160.
- Mente-cuerpo: 18, 26,
- Joseph Capgras: 132, 133, 134, 135, 136, 138, 148.
- Kepler: 19.
- Kuhn: 103.
- Lakoff: 140.
- Lenin: 28.
- Ludwig Wittgenstein: 43, 100.
- Malebranche: 22.
- Marr: 68, 69, 80.
- Michaels: 80.
- Michel Foucault: 103, 119.
- Noam Chomsky: 39, 87.
- Otto Neurath: 29.
- Palatinus: 80.
- Paul Churchland: 36, 37, 60, 98.
- Paul Thagard: 55, 64.
- Pierre Gassendi: 27.
- Poggio: 68, 69, 80.
- Rene Descartes: 21, 22, 27, 71, 89, 105, 117, 143, 155, 161.
- Rudolf Carnap: 29, 43.
- Searle: 60, 61, 62, 66, 104.
- Shapiro: 74, 75.
- Sinigaglia: 98.
- Skinner: 39, 40.
- Thomas Hobbes: 27.
- Thomas Nagel: 67, 73, 148, 153, 154.
- Thompson: 83, 85.

28, 29, 44, 71, 73, 153.

- Microelectrodos: 99.
- Monos títes: 97.
- NC (neuronas canónicas): 95.
- NE (Neuronas espejo): 38, 77, 94, 111, 124.
- Neurociencias: 14, 17, 36, 53, 59, 116.
- Neuronas: 34, 35, 36, 37, 38, 63, 93, 94, 95, 96, 98, 99, 111, 116.
- Nichos ecológicos: 14.
- Ocasionalismo: 22.
- OMS: 121.
- Patología: 18, 74, 114, 120, 121, 122, 134, 135, 136, 148, 151, 152.
- Psicología ecológica: 17, 77, 78, 79, 84, 87, 90, 95.
- Psicopatología: 10, 17, 18, 116, 118, 119, 120, 159, 160.
- Psiquiatría: 119.
- Regiones cerebrales: 97.
- Representación: 23, 24, 56, 61, 62, 63, 64, 65, 68, 72, 79, 87, 125.
- Resonancia: 97, 98, 137.
- Síndrome: 121, 122, 134.
- Sociología: 97, 103, 110.
- Subjetividad: 108, 117, 136.

• Zahavi: 147.

• Zenon Pylyshyn: 57.

- superveniencia: 30, 31, 33, 38.
- TCRM: 24, 53, 54, 55, 56, 69, 70, 71, 72, 76, 77, 78, 79, 83, 84, 100, 105, 159.
- teoría computacional: 24, 52 53, 70.
- Tomografías: 99.
- Trascendental: 117.
- Trastorno: 122, 124, 125, 126, 127, 131, 134, 135, 136, 147.
- TS (teoría de la simulación): 106.
- TT (teoría-teoría): 106.

SUMARIO

9	Presentación
13	Introducción
19	I. La mente aislada
24	1. Filosofía de la mente: explicando por reducción
26	<i>Fisicismo: la mente es cerebro</i>
37	<i>Conductismo: la mente es conducta</i>
45	<i>Funcionalismo: la mente es una función</i>
52	2. La perspectiva clásica de la mente
57	<i>La tesis computacional</i>
62	<i>La tesis de la representación</i>
71	II. Trascender la mente
77	1. Organismo y medio: la psicología ecológica
82	2. Enacción: mente, cuerpo y mundo
89	3. Las extensiones de la mente
93	4. Interludio: neuronas y el espejo de la intersubjetividad
101	5. Comprensión, cultura y empatía

115	III. La mente deficiente
119	1. Filosofía de la psicopatología
125	2. Delirios
127	<i>Retos filosóficos del delirio</i>
131	3. Reconocimiento del mundo
136	<i>Retos filosóficos en el reconocimiento del mundo</i>
141	4. Cuerpo, propiedad y agencia
143	<i>Retos filosóficos sobre el cuerpo</i>
147	5. Conciencia
152	<i>Retos filosóficos sobre las patologías de la conciencia</i>
159	Conclusiones
163	Referencias
175	Índices terminológico y onomástico
181	Sumario

*La mente en el mundo. Introducción al
proyecto de la cognición corporizada*

se terminó de imprimir en marzo de 2022,
en los talleres de la Universidad Autónoma de la Ciudad de México,
San Lorenzo, 290, col. Del Valle, alcaldía Benito Juárez, C.P. 03100, Ciudad de
México. El tiraje fue de 500 ejemplares.
Cuidado de la edición: Ángeles Godínez Guevara
Diseño editorial: Sergio Cortés Becerril



Adrián Espinosa Barrios es licenciado en filosofía por la Universidad de Guadalajara. Estudió la maestría y el doctorado en filosofía de la ciencia con especialidad en filosofía de las ciencias cognitivas en el Instituto de Investigaciones Filosóficas de la UNAM. Es parte del cuerpo académico del Posgrado en Ciencias Cognitivas de la Universidad Autónoma de Morelos.

Actualmente es profesor de tiempo completo de la Academia de Filosofía e Historia de las Ideas en la UACM-Cuauhtepc en donde ha publicado el libro *Temas de Cultura Científica y Tecnológica. Actitud Crítica y Ciudadanía Responsable*. Su tema de interés es la filosofía de la conciencia en intersección con la fenomenología husserliana.

El lector tiene en sus manos un texto que es principalmente para las y los estudiantes de la carrera de Filosofía e Historia de las Ideas de la UACM. Pretende compilar la información necesaria sirviendo de guía de lectura a los textos de investigación que se requieren para seguir la formación en el área. Se estructura en tres partes: en la primera el estudiante se adentra a la discusión del problema mente cuerpo por medio del repaso de las principales corrientes filosóficas de corte cognitivista que dominaron la primera mitad del siglo XX; en la segunda parte se expone el tema de las teorías de la corporización que son aquellas que surgen como crítica a la concepción computacional de la mente y que hoy en día forman una parte esencial del estudio filosófico de la mente y la conciencia; finalmente, el tercer capítulo presenta al estudiante una nueva perspectiva de estudio filosófico de la mente: la filosofía de la psicopatología. De esta manera, el texto ha de servir al mismo tiempo como introducción histórica al área, como presentación de los temas de interés actual y como argumentación de la necesidad de llevar a cabo una filosofía científicamente informada en todas las áreas, pero muy especialmente, en lo que concierne al estudio de la mente y la conciencia.