



EL INSTINTO DE LA CONCIENCIA



MICHAEL S. GAZZANIGA



CÓMO EL CEREBRO CREA LA MENTE

PAIDÓS

Índice

Portada

Sinopsis

Portadilla

Dedicatoria

Introducción

Parte I. El camino hacia el pensamiento moderno

1. La rígida, inestable y necia concepción histórica de la conciencia
2. Los albores del pensamiento empírico en filosofía
3. Los grandes progresos del siglo XX y los inicios del pensamiento moderno

moderno

Parte II. El sistema físico

4. Cómo fabricar cerebros módulo a módulo
5. Empezamos a entender la arquitectura cerebral
6. El abuelo está demente, pero conserva la conciencia

Parte III. Llega la conciencia

7. El concepto de complementariedad: el regalo de la física
8. De lo no viviente a lo viviente y de las neuronas a la mente
9. Torrentes de burbujas y conciencia personal
10. La conciencia es un instinto

Agradecimientos

Notas

Créditos

Gracias por adquirir este eBook

Visita Planetadelibros.com y descubre una nueva forma de disfrutar de la lectura

¡Regístrate y accede a contenidos exclusivos!

Primeros capítulos

Fragmentos de próximas publicaciones

Clubs de lectura con los autores

Concursos, sorteos y promociones

Participa en presentaciones de libros

PlanetadeLibros

Comparte tu opinión en la ficha del libro
y en nuestras redes sociales:



Explora

Descubre

Compa

Sinopsis

La idea del cerebro como máquina, propuesta por primera vez hace siglos, ha llevado a suposiciones sobre la relación entre la mente y el cerebro que los científicos y los filósofos persiguen hasta nuestros días. Gazzaniga, sin embargo, afirma que se trata precisamente de lo contrario: los cerebros hacen máquinas, pero no pueden reducirse a una. Una nueva investigación sugiere que el cerebro es en realidad una confederación de módulos independientes que trabajan juntos. Comprender cómo la conciencia podría emanar de tal organización ayudará a definir el futuro de la neurociencia y la inteligencia artificial, y cerrará la brecha entre el cerebro y la mente.

EL INSTINTO DE LA CONCIENCIA

Cómo el cerebro crea la mente

Michael S. Gazzaniga

Traducción de Francisco J. Ramos

PAIDÓS 
Barcelona • Buenos Aires • México

*Para Leonardo,
una conciencia en desarrollo donde las haya*

Introducción

Imagine, si puede, que únicamente fuera consciente de un solo momento: el ahora. Ese momento existe sin pasado ni futuro. Ahora imagine que la vida fuera una serie de esos momentos, cada uno de los cuales existe en una especie de momento aislado con respecto a todos los demás, no conectados por el tiempo subjetivo. Imagínese temporalmente congelado en cada uno de los momentos que juntos forman una vida normal. Resulta difícil concebir ese escenario porque nuestra mente viaja de aquí para allá a través del tiempo con tanta soltura como una bailarina en *El cascanueces*. Un momento sirve como materia prima para la siguiente acción planificada, que a su vez se evalúa en el presente comparándola con nuestra experiencia pasada. Es difícil imaginar que pueda no ser así; sin embargo, basta con darse un golpe en la cabeza de la manera apropiada y usted mismo podría encontrarse en esa situación: todavía capaz de entender el concepto de tener un pasado y un futuro, pero incapaz de ubicarse en un pasado o futuro propios. Extraño, pero cierto. Sin pasado; sin futuro: solo el presente.

En este libro llevaré al lector de viaje a un mundo donde son habituales una serie de alteraciones de lo que denominamos «experiencia consciente», por difíciles de imaginar que resulten. La planta de neurología de cualquier hospital está repleta de trastornos de la experiencia consciente normal. Cada uno de esos casos nos dice algo acerca de cómo se organizan nuestros cerebros para proveernos de nuestra preciada conciencia, momento a momento. Cada ejemplo de alteración pide a gritos ser descifrado, ser utilizado para deducir una historia coherente acerca de cómo nuestro cerebro construye y produce las alegrías cotidianas que implica el hecho de ser

conscientes. En el pasado, la gente se contentaba con contar historias sobre esos extraños fenómenos; hoy, en el siglo XXI, no basta con limitarse a describir toda esa variedad de intrigantes trastornos. Mi objetivo en este libro es avanzar en la investigación del problema de la conciencia, y trataré de ilustrar cómo nuestro cerebro —un órgano exquisitamente evolucionado— se organiza para hacer su magia. En suma: pretendo examinar cómo la materia crea la mente.

Hace unos años, en un viaje de trabajo, tuve que pasar la aduana en el aeropuerto londinense de Heathrow. El encargado del control de pasaportes, un atento funcionario británico, me preguntó diligentemente por mi nombre y profesión, y la razón por la que viajaba al Reino Unido. Yo le dije que me dedicaba a investigar sobre el cerebro y que iba a Oxford a una reunión. Me preguntó si conocía las diferentes funciones de los dos hemisferios cerebrales, el izquierdo y el derecho. No sin cierto orgullo, le respondí que no solo había oído hablar de ello, sino que en parte era responsable del trabajo que lo había dado a conocer. Mientras examinaba mi pasaporte, me preguntó de qué iba a tratar la reunión en Oxford. Le contesté, con aire de autoridad:

—De la conciencia.

El funcionario cerró el pasaporte y, al devolvérmelo, inquirió:

—¿Alguna vez ha pensado en dejarlo correr?

Aparentemente no. Algunos de nosotros somos naturalmente incautos en nuestro deseo de preguntarnos por nuestra propia naturaleza. Trabajar en la ciencia de la mente y del cerebro como yo llevo haciendo sesenta años me permite tener claro que los humanos todavía no hemos comprendido el problema en toda su plenitud. Aun así, está en nuestra naturaleza pensar en quiénes y qué somos y en qué significa ser conscientes. Una vez nos pica la curiosidad sobre esta cuestión, nos pasamos la vida corroídos por el deseo de encontrar una respuesta. Sin embargo, cuando intentamos aferrar el problema de la conciencia, este parece disolverse como la niebla. ¿Por qué el esfuerzo por entender la conciencia se ha revelado tan difícil? ¿Acaso las persistentes

ideas del pasado nos impiden ver claramente cómo se produce? ¿Es la conciencia solo lo que hace el cerebro? Al igual que un reloj de bolsillo, con sus engranajes, nos da la hora, ¿el cerebro, con sus neuronas, nos da la conciencia? La historia del tema es extensa y se ha visto sacudida por oscilaciones pendulares entre los mecanicistas puros y los esperanzados mentalistas. Sorprendentemente, dos mil quinientos años de historia humana no han resuelto la cuestión ni han enseñado a nuestra especie a enmarcar la comprensión de nuestra personal experiencia consciente. De hecho, nuestras ideas básicas al respecto no han cambiado tanto. Aunque fue Descartes quien dio origen al pensamiento sobre la conciencia hace trescientos años, ha habido dos concepciones generales contradictorias —que la mente forma parte del funcionamiento del cerebro, o, por el contrario, que de algún modo funciona de manera independiente de este— que parecen haber existido desde siempre. De hecho, estas ideas todavía nos acompañan.

En los últimos años el tema de la conciencia se ha convertido de nuevo en una cuestión candente. Al mismo tiempo, y a pesar de la actual avalancha de nuevos datos, hay muy pocas propuestas que cuenten con una aceptación generalizada —por no decir ninguna— acerca de cómo el cerebro construye una mente y, con ella, una experiencia consciente. El objetivo del presente volumen es salir de este atolladero y, a la vez, presentar una nueva visión de cómo conceptualizar la conciencia. Nuestro viaje incluye el conocimiento obtenido no solo a partir de la neurología, sino también de la biología evolutiva y teórica, la ingeniería y la física, y, por supuesto, la psicología y la filosofía. Nadie dijo que la búsqueda iba a ser sencilla, pero la meta —comprender cómo la naturaleza realiza el truco de convertir las neuronas en mentes— es perfectamente alcanzable. Así que ¡allá vamos!

Dicho lisa y llanamente: yo creo que la conciencia es un instinto. Muchos organismos —no solo los humanos— ya vienen con ella incorporada. Eso es lo que son los instintos: algo que los organismos llevan incorporado. Los seres vivos —aunque estén hechos de los mismos materiales que el mundo natural

no viviente que los rodea— poseen una organización que permite que exista la vida y, en última instancia, la conciencia. Y los instintos abarcan a todos los organismos, desde las bacterias hasta los humanos. La supervivencia, el sexo, la resiliencia y la marcha normalmente se consideran instintos, pero también lo son otras capacidades más complejas como el lenguaje y la sociabilidad: todas son instintos. La lista es larga, y parece que los humanos tenemos más instintos que otras criaturas; sin embargo, el instinto de la conciencia tiene algo especial. No es un instinto ordinario. De hecho, parece tan extraordinario que muchos creen que solo los humanos podemos atribuirnoslo. Aunque ese no es el caso, queremos saber más sobre él. Y, dado que todos lo tenemos, todos pensamos que tenemos alguna idea al respecto, pero, como veremos, se trata de un instinto resbaladizo y complejo, situado en el órgano más impenetrable del universo: el cerebro.

La palabra *manzana* es un sustantivo; denota un objeto físico real. La palabra *democracia* también es un sustantivo, aunque describe algo un poco más difícil de precisar, un estado de relaciones sociales. Es fácil para mí mostrarle una manzana, su realidad física; en cambio, me resulta difícil mostrarle la realidad física de una democracia. ¿Y en el caso de *instinto*, que es otro sustantivo? Los tres son «algunos» definibles, ya sean objetos o conceptos, que maneja el cerebro. Tenemos muchos de esos «algunos», pero ¿dónde residen en el cerebro? ¿Es acertado representar algunos de ellos como estructuras cerebrales reales y otros como acciones procesuales de determinadas estructuras? De hecho, ¿cuál es la realidad física de un instinto? ¿Es tangible, como una manzana, o esquiva, como una democracia?

Los instintos complejos se parecen más a las democracias; son identificables, pero no fácilmente localizables. Surgen de la interacción de instintos simples, pero no lo son en sí mismos, de modo similar a como un complejo reloj de bolsillo se las arregla para llevar la cuenta del tiempo, pero resulta imposible encontrar el tiempo en sí en un reloj. Para entender cómo el reloj se relaciona con el tiempo, hay que describir cómo está diseñado, cuál es

su arquitectura; no basta con enumerar todos sus resortes y engranajes. Lo mismo vale para el instinto de la conciencia. No crea que por el mero hecho de que la conciencia sea un instinto tiene que haber una red cerebral única, unitaria y discreta que genere ese extraordinario estado de autoconocimiento del que todos disfrutamos. No es ese el caso en absoluto. Si, armados con nuestras ideas, nos dedicamos a visitar plantas de neurología, reconoceremos de inmediato que los pacientes que sufren demencia, incluso demencias graves, son conscientes. Esos pacientes, con una vastísima variedad de lesiones cerebrales, con un nivel de alteración lo suficientemente amplio como para derrotar a cualquier ordenador, siguen siendo conscientes. En una habitación de hospital tras otra —cada una de ellas ocupada por un paciente que sufre una disfunción cerebral focal o difusa— puede percibirse el murmullo de la conciencia. Tras un recorrido por diversas plantas, empieza a parecer que la conciencia no es en absoluto una propiedad del sistema, sino que depende de circuitos cerebrales locales.

En la primera parte de este libro veremos cómo la naturaleza se convirtió en un «eso», algo independiente de nosotros que puede estudiarse y entenderse en términos objetivos. Seguiremos el hilo de esta idea a través de Descartes hasta los tiempos modernos y los albores de la biología contemporánea. Sorprendentemente, la mayor parte del pensamiento científico moderno ha retrocedido para pasar a basarse en las ideas de los antiguos griegos, y sustenta esencialmente los mismos modelos, que vinculan de forma inexorable lo mental y lo físico en un único sistema. La ciencia contemporánea ha empezado a perseguir el mismo objetivo que buscaban los griegos, pero hasta ahora se ha quedado corta. Una vez más, hacen falta nuevas ideas, y este libro pretende dar un paso en ese sentido.

En la segunda parte se introducen algunos principios modernos del funcionamiento del cerebro que creo que deberían guiar nuestra exploración para entender cómo las neuronas producen mentes. No deja de sorprenderme el hecho de que la metáfora del «cerebro como máquina», propuesta

inicialmente por Descartes y adoptada plenamente por la mayoría de la ciencia moderna, nos haya llevado a creer que se necesita la máquina entera para llevar a cabo muchas de sus funciones. En realidad, cada uno de nosotros es una confederación de módulos bastante independientes, orquestados para trabajar juntos. Para entender cómo dichos módulos colaboran entre sí, necesitamos conocer la arquitectura global del sistema, una arquitectura que recibe el nombre de «estratificación» (*layering*) y que sin duda resultará familiar a muchos lectores, como, por ejemplo, a los informáticos. Por último, visitaremos la clínica neurológica para poner a prueba esa formulación. Allí descubriremos que nuestro cerebro modular, con su arquitectura estratificada, está gestionando nuestra conciencia desde... todos y cada uno de sus tejidos locales, constantemente. No hay un único sistema centralizado que trabaje para producir la grandiosa magia de la experiencia consciente, sino que esta está en todas partes, y parece que no se puede eliminar, ni siquiera con una enfermedad cerebral de gran alcance como la enfermedad de Alzheimer.

En la tercera parte del libro afronto esa persistente cuestión que constituye el núcleo del tema de la relación mente/cerebro. ¿Cómo las neuronas informan a la mente? ¿Cómo esos blandos manojos de tejido húmedo nos convierten a usted y a mí en seres mentales? Resulta que hay brechas en nuestra comprensión del mundo físico. Estudiamos un nivel de organización y luego otro, pero en realidad no entendemos cómo esos dos niveles distintos trabajan juntos. Existe una notoria brecha entre la vida y la materia inanimada, entre la mente y el cerebro, entre el mundo cuántico y nuestro mundo cotidiano. ¿Cómo se pueden salvar esas brechas? Me parece que aquí la física puede ayudarnos.

Por último, ofrezco una perspectiva de cómo esos módulos, capas y brechas se combinan para producir lo que denominamos «experiencia consciente». El profesor de psicología Richard Aslin me comentaba en cierta ocasión que él creía que la idea de «conciencia» era un sustituto a toda una serie de variables relacionadas con nuestra vida mental. Utilizamos «conciencia» como una forma abreviada de describir fácilmente las funciones

de múltiples mecanismos innatos e instintivos como el lenguaje, la percepción y la emoción. Resulta evidente que la conciencia se entiende mejor si la conceptualizamos también como un instinto complejo. Todos nosotros nacemos con un montón de instintos. Nuestra incesante pauta de pensamiento salta constantemente de una cosa a otra. Tenemos determinadas sensaciones en torno a una idea, luego la contraria, luego nuestra familia, luego una comezón, luego una melodía favorita, luego la próxima reunión, luego la lista de la compra, luego ese colega irritante, luego nuestro equipo de fútbol, luego... Y así sucesivamente hasta que aprendemos, casi en contra de nuestro ser natural, a tener un pensamiento lineal.

El pensamiento lineal consciente es una ardua tarea. Ahora mismo estoy sudando. Es como si nuestra mente fuera una burbujeante olla de agua hirviendo: es difícil predecir qué burbuja llegará a la superficie en un momento dado. La burbuja que finalmente lo consigue estalla dando origen a una idea, pero solo para ser reemplazada de inmediato por otras burbujas. La superficie está permanentemente pletórica de actividad, de interminable actividad, hasta que las burbujas se van a dormir. La flecha del tiempo las vincula unas a otras aunque cada una de ellas surge en su momento. Considere la posibilidad de que tal vez —solo tal vez— la conciencia solo pueda entenderse como las burbujas del cerebro, cada una con su propio hardware para salvar la brecha y tener su momento. Si esta imagen le resulta abstrusa, lea el libro para descubrir por sí mismo si al final también logra verlo igual. Y lo que es más importante: disfrute de las burbujas de sus pensamientos a medida que estas emerjan a la superficie de su propia conciencia.

PARTE I

El camino hacia el pensamiento moderno

La rígida, inestable y necia concepción histórica de la conciencia

—¡Habla en cristiano! —protestó el Aguilucho—. No sé lo que quieren decir ni la mitad de estas palabras altisonantes, y es más, ¡creo que tampoco tú sabes lo que significan!

LEWIS CARROLL, *Alicia en el País de las Maravillas*

Sigmund Freud murió el año en que yo nací, 1939. Ese año circularon un montón de ideas disparatadas sobre la naturaleza de nuestra vida psíquica, muchas de ellas concebidas por el propio Freud. Popularmente no se le recuerda como tal, pero Freud era en el fondo un biólogo, un biólogo reduccionista. Estaba comprometido con la creencia de que el cerebro generaba la mente de una forma determinista, una visión compartida por muchos de los neurocientíficos actuales. Hoy reconocemos que muchas de sus ideas eran pura fantasía, pero hasta la década de 1950 fueron tan ampliamente aceptadas que llegaron a convertirse en el testimonio dominante en todas las cuestiones psicológicas dirimidas por un tribunal estadounidense.

Ha sido en mi época, no en la de Freud, cuando la humanidad ha aprendido más acerca de cómo el cerebro realiza sus trucos. La especulación descabellada sobre las fuerzas que gobiernan nuestra vida mental ha dado paso al conocimiento concreto sobre las influencias moleculares, celulares y ambientales que subyacen a nuestra existencia. De hecho, los últimos setenta y cinco años de investigación han proporcionado una gran cantidad de

información sobre el cerebro, a veces revelando incluso principios organizativos. Estoy seguro de que Freud se habría deleitado en nuestro nuevo mundo y habría estado encantado de poner a trabajar su increíble imaginación en la nueva ciencia del cerebro. Sin embargo, los profundos enigmas que en el siglo pasado enfrentaron a científicos de toda laya —y que, de hecho, se remontan a los antiguos griegos— siguen estando presentes en la actualidad. ¿Cómo demonios la materia inanimada se convierte en el componente básico de los seres vivos? ¿Cómo las neuronas devienen mentes? ¿Qué vocabulario habría que emplear para describir las interacciones entre el cerebro y su mente? Cuando la humanidad obtenga algunas de las respuestas a esas preguntas, ¿nos sentiremos decepcionados por ellas? ¿Nuestra futura comprensión de la conciencia nos resultará simplemente insatisfactoria? ¿Será sencilla, pero a la vez fría y rigurosa?

Introducirse en la historia del estudio de la conciencia resulta desalentador. Por una parte, está plagada de textos complejos y abstractos escritos por filósofos. Incluso John Searle, uno de los principales filósofos actuales de la conciencia, ha admitido: «Probablemente debería leer más filosofía de la que leo. Pero creo que muchas obras de filosofía son como una endodoncia: solo piensas que tienes que pasar ese mal trago». ¹ Añádase a ello la opinión del gran filósofo David Hume, que aportaba argumentos contundentes aduciendo que la mayoría de las preguntas planteadas por los filósofos simplemente no podían responderse utilizando las metodologías de la lógica, las matemáticas y la razón pura. Aun así, los filósofos nos han hecho pensar sobre la mente, el alma y la conciencia; y desde tiempos antiguos han tenido una enorme influencia.

El concepto «conciencia» es relativamente moderno. La propia palabra, tal como hoy se utiliza de forma generalizada en docenas de contextos distintos (Marvin Minsky la llamaría una «palabra maleta» porque está llena de significados diversos), solo adquirió su sentido moderno a mediados del siglo XVII con René Descartes. Tiene su origen en el término griego *oida* —«haber

visto» o «percibido» y, por ende, «saber»— y su equivalente latino *scio*, «saber». Pero los antiguos no tenían un concepto explícito de conciencia. Les interesaba saber cómo funcionaba la mente, de dónde venían los pensamientos e incluso si ello implicaba un proceso puramente físico, pero la mayoría de los antiguos pensadores acabaron llegando a la conclusión de que la vida mental era el producto de un espíritu inmaterial. Y, cuando la conciencia se define como un espíritu inmaterial, se hace difícil empezar a pensar en los mecanismos subyacentes.

A lo largo de los siglos, el concepto de «mente» y el concepto de «alma» se han visto involucrados en una relación intermitente. Durante la mayor parte de la historia escrita, la mera idea de que la realidad psíquica personal fuera una «cosa», algo susceptible de ser estudiado, fue prácticamente inexistente. Presumiblemente, nuestro cerebro, nuestras estructuras de pensamiento y nuestras emociones no han cambiado, así que, ¿qué era lo que pensábamos los humanos? Sin embargo, como quedará de manifiesto, el concepto de «conciencia» ha cambiado de manera radical en los últimos veinticinco años: sus etéreos comienzos tienen muy poco que ver con su significado actual.

Los humanos necesitamos una nueva forma de concebir el problema, y, con un poco de suerte, este libro puede ofrecer algunos puntos de partida nuevos. Sin embargo —y como ocurre siempre—, es conveniente mirar hacia atrás antes de lanzarse hacia delante.

Los primeros pasos: aciertos y errores

Los antiguos egipcios y mesopotámicos fueron los precursores filosóficos del mundo occidental. En su concepción del mundo, la naturaleza no era un adversario en la lucha por la vida. Antes bien, hombre y naturaleza viajaban en el mismo barco, eran compañeros en una misma aventura. El hombre concebía el mundo natural en los mismos términos en los que se concebía a sí mismo y a los demás. El mundo natural tenía pensamientos, deseos y

emociones, exactamente igual que los humanos. Por lo tanto, el reino del hombre y el de la naturaleza eran indistinguibles y no tenían que entenderse de formas cognitivamente distintas. Los fenómenos naturales se concebían en los mismos términos que la experiencia humana: generosos o no tanto, fiables o malévolos, etc. Los antiguos pensadores de Oriente Próximo sí reconocían la relación entre causa y efecto, pero a la hora de especular sobre ello adoptaron la perspectiva de atribuir el origen a un «quién» antes que a un «qué». Cuando crecía el Nilo, era porque el río quería hacerlo, no porque hubiera llovido. No existía un conocimiento científico que sugiriera lo contrario.

No fue ese el caso de los antiguos griegos. Los primeros filósofos helenos no eran sacerdotes que hubieran recibido el encargo de gestionar los asuntos espirituales por parte de sus respectivas comunidades, como ocurría en Oriente Próximo. No eran videntes profesionales. Eran un grupo de aficionados que se dedicaban a holgazanear libres de las restricciones del dogma, espoleados por su curiosidad hacia el mundo natural y encantados de compartir sus pensamientos. Cuando empezaron a preguntarse por sus orígenes, no se preguntaron «quién» era el progenitor, sino «qué» fue la causa primera, lo cual supuso un cambio de perspectiva monumental para la humanidad, que el arqueólogo y egiptólogo Henri Frankfort califica de «impresionante»:

Aquellos hombres partieron, con absurda audacia, de un supuesto no verificado en absoluto. Sostenían que el universo es un todo inteligible. En otras palabras, suponían que un orden único subyace al caos de nuestras percepciones y que, además, podemos llegar a comprender ese orden.²

Frankfort pasa luego a explicar cómo los filósofos griegos pudieron dar ese salto:

La diferencia fundamental entre la actitud del hombre moderno y la del antiguo con respecto al mundo circundante es la siguiente: para el hombre científico moderno, el mundo fenoménico es básicamente un «Eso»; para el hombre antiguo —como para el primitivo— es un «Tú».

Un «Tú» es alguien con creencias, pensamientos y deseos, que actúa por sí mismo y, por tanto, no es necesariamente estable o predecible. Por otro lado, «Eso» es un objeto, no un amigo. «Eso» puede relacionarse con otros objetos en lo que puede parecer la más razonable de las organizaciones. Partiendo de dichas relaciones, es posible expandirlas y buscar leyes universales que gobiernen el comportamiento y los acontecimientos en función de condiciones predecibles y prescritas. Buscar la identidad de un objeto es un proceso activo. Por el contrario, entender a un «Tú» es un proceso pasivo en el que uno empieza por recibir una impresión emocionalmente sesgada. Un «Tú» es único e impredecible, y solo se lo conoce en la medida en que se revela. Cada experiencia del «Tú» es individual. De una interacción con un «Tú» puedes obtener una historia o un mito, pero no puedes extraer una hipótesis. La transición del «Tú» al «Eso» hizo posible el pensamiento científico.

Este enorme salto de perspectiva de los griegos creó la atmósfera que catapultaría a Aristóteles hacia una vida científica. La postura de Aristóteles era que la labor de la ciencia consistía en dar cuenta objetivamente del «por qué» de las cosas, lo que condujo a su doctrina de la causalidad. Para él, el conocimiento científico sobre algo (pongamos X) incluía todas las posibles formas de responder a la pregunta de «por qué»: si X estaba causado por Y, o si Y era al menos una condición necesaria para que ocurriera X, ese era el tipo de afirmación que correspondía a la ciencia. Postuló cuatro categorías causales: material, formal, eficiente y final. Por ejemplo, si alguien le preguntara «Aristóteles, ¿por qué un carro?», él respondería que la causa material era la madera, que la causa formal era el plano de construcción, que la causa eficiente era su construcción y que la causa final era... simplemente que quería tener uno.

Para Aristóteles, el mundo natural era un entramado de lo que el teórico biológico Robert Rosen denomina implicaciones causales: X conlleva todas sus Y (en plural). Rosen señala que la idea de Aristóteles era mostrar que ninguna forma de explicación basta por sí sola para entender algo, dado que

las categorías causales no se implican mutuamente. Por ejemplo, saber cómo construir algo no implica entender cómo funciona; y viceversa: saber cómo funciona algo no implica saber construirlo. Además, para Aristóteles, lo que determinaba la ciencia era el contenido, que era independiente del método por el que se estudiaba.

El método científico tal como hoy se practica es un sistema formal en el que una hipótesis produce sus inferencias, esto es, sus efectos: la hipótesis implica dichos efectos. Otra forma de decirlo es que la causa precede al efecto, lo cual plantea un problema al formular la cuestión de la causa final del «por qué» de Aristóteles. Volvamos a «¿Por qué el carro, Aristóteles?». ¿Por qué Aristóteles tenía un carro estacionado frente a su casa cuando horas antes había estado estacionado en Almacenes Acrópolis? Él había visto el carro (lo que implicaba los efectos de las causas material, formal y eficiente) y quería tenerlo. Aquí cambiaban las tornas y el efecto precedía a la causa. Es una doble negación en el mundo newtoniano, donde un estado solo puede implicar estados subsecuentes. En consecuencia, la causalidad final de Aristóteles, como categoría independiente, desapareció de la escena científica. Más adelante veremos el daño que esto ha causado a la biología.

Entre otras cosas, Aristóteles quería saber más sobre el cuerpo humano y sobre cómo funcionaba. No era tarea fácil, ya que para los griegos la disección humana era un tabú. Aristóteles esquivó el problema realizando numerosas disecciones de animales. Partiendo de lo aprendido, ideó un sistema de clasificación de los organismos, la *scala naturae*, una escala jerárquica graduada basada en el tipo de «alma» que cada uno de ellos poseía. En la base estaban las plantas, que él postuló que tenían un «alma vegetativa» responsable de su crecimiento y reproducción. Ni que decir tiene que el hombre se situaba en la parte superior de aquella *scala naturae*.

Pero Aristóteles no se detuvo ahí. También propuso que los animales poseían un «alma sensible» que impulsaba el movimiento autónomo, la percepción, la sensación, el apetito y la emoción. En el caso único de los

humanos, dentro del alma sensible residía un «alma racional» que nos dotaba de las especiales facultades de la razón, la voluntad racional, el pensamiento y la reflexión, a la par que nos diferenciaba de todos los seres situados por debajo en la *scala*. Pero la idea más importante, que refleja la revolución acaecida en el pensamiento humano, era que el «conocimiento» de aquellas facultades llegaba no por pura introspección o divagaciones mentales, sino observando cómo uno está conectado con el mundo circundante. El «Eso», es decir, un objeto como el mundo que nos rodea, podía ser estudiado y examinado. Solemos olvidar que esta sencilla idea, hoy comúnmente aceptada, no existía hace unos miles de años. Es evidente que las ideas tienen consecuencias, y, felizmente, seguimos sintiéndonos cautivados por la idea y el poder de la observación científica.

Aristóteles acertó al entender el proceso de la ciencia, pero erró en todas sus conclusiones sobre el origen de los pensamientos. Si un estudiante moderno cometiera un error como el que él cometió, le suspenderían sin ninguna duda. Aristóteles sabía, por las acciones de los animales y los humanos, que estos podían percibir el mundo; y en sus disecciones observó que algunos animales carecían por completo de un cerebro visible. En consecuencia, concluyó que aparentemente el cerebro no contaba demasiado. Lo primero que veía aparecer en los embriones que estudió era el corazón, de manera que situó allí el alma, que en el caso de los humanos incluía el alma racional. Para él, el concepto de «alma» no tenía un significado espiritual, puesto que no creía que esta sobreviviera a la muerte, sino que hacía referencia al órgano que daba lugar a la sensación, a nuestro conocimiento del mundo. Pensaba que el alma racional, que era la fuente del intelecto humano, requería ciertos mecanismos perceptivos; en consecuencia, requería un cuerpo con sus partes y órganos. Pero no creía que hubiera una parte del cuerpo o un órgano que pensara. Aristóteles ni siquiera musitó la palabra *consciente*, pero sí se preguntó: «¿Cómo conocemos nuestras propias percepciones?». En

términos generales, diremos que Aristóteles fue el que consiguió que la gente empezara a pensar en la naturaleza física de la humanidad.

Los enormes avances iniciados en Grecia se exportaron rápidamente. En el año 322 a.C., no mucho después de la muerte de Aristóteles, Herófilo y Erasítrato, dos médicos griegos que vivían en Alejandría, decidieron desafiar el tabú de diseccionar cuerpos humanos y se pusieron a ello, convirtiéndose en los primeros en descubrir el sistema nervioso y en escribir sobre él. También descubrieron los ventrículos, las cámaras vacías que alberga el cerebro. Herófilo decidió que aquellas cámaras debían de ser la sede del intelecto, y que desde ellas fluían espíritus que viajaban a través de los nervios huecos hasta los músculos, haciendo que estos se movieran. Aunque su interpretación no fuera del todo exacta, en general se les atribuye el mérito de ser los primeros neurocientíficos. Por increíble que hoy pueda parecernos, la cultura griega que diseñó y construyó el Partenón no conocía el cerebro; y la cultura egipcia que diseñó y construyó las pirámides no tenía ni la menor idea de su funcionamiento.

La historia siguió traqueteando durante otros cuatrocientos años, un microsegundo en el tiempo evolutivo. Roma se convirtió en la fuerza dominante en el Mediterráneo y de algún modo logró atraer al extraordinario médico Claudio Galeno, originario de Pérgamo, una ciudad griega situada en la costa egea de la actual Turquía. Galeno completó su formación médica como empirista tras haberse sumergido en las enseñanzas de Herófilo y Erasítrato en Alejandría, entonces bajo el dominio romano. En la antigua Grecia, la escuela empírica de la práctica médica se basaba en la observación de los fenómenos y en la experiencia, no en dictámenes dogmáticos. Galeno regresó a Pérgamo para desempeñar su primer trabajo: médico de gladiadores. Dado que los romanos, como los griegos, no permitían la disección humana, Galeno nunca realizó ninguna, pero perfeccionó su conocimiento de la anatomía y la cirugía con los ensangrentados restos de sus pacientes y practicando diariamente disecciones de animales, sobre todo monos de

Berbería. Aunando sus conocimientos de primera mano, la saludable ayuda de las enseñanzas de sus lejanos mentores —Herófilo y Erasístrato— y una pizca de la teoría de Hipócrates de que el cuerpo estaba compuesto por cuatro humores, ideó una nueva concepción del cuerpo y su funcionamiento, lo que le valió una reputación estelar. No tardó en desplazarse a Roma, y su creciente fama le llevó a convertirse en el médico personal del emperador, Marco Aurelio.

Las contribuciones de Galeno a la medicina son impresionantes. Fue el primero en percibir que existía una diferencia entre la sangre arterial y la venosa. Hoy sabemos que la sangre arterial es rica en oxígeno, mientras que la sangre venosa transporta mucho menos oxígeno (los tejidos se lo han robado para poder respirar); una diferencia de la que sacamos partido en el estudio del cerebro mediante imagen por resonancia magnética funcional (IRMf), que constituye la piedra angular de la neurociencia moderna. Galeno fue el primero en describir las cuatro cámaras del corazón humano, y actualizó el conocimiento de los sistemas circulatorio, respiratorio y nervioso. Desde luego, también cometió algunos errores anatómicos, uno de los cuales fue afirmar la existencia de una red de vasos sanguíneos —la *rete mirabile*— que situó en la base del cráneo humano basándose en sus disecciones de bueyes. Fue un error importante, que a la vez constituye una advertencia sobre el razonamiento inductivo. Como se demostraría años después, los humanos no tenemos *rete mirabile* en absoluto.

No obstante, Galeno entendió que el alimento y la respiración eran necesarios para la vida humana, sosteniendo que el cuerpo los transformaba en carne y espíritu. Combinando las obras de Hipócrates, Platón, Sócrates y Aristóteles, ideó el concepto de un alma material tripartita. Partiendo de la descripción platónica de las almas racional, espiritual y apetitiva, asignó a cada una de ellas una ubicación anatómica: el alma racional residía en el cerebro, el alma espiritual en el corazón y el alma apetitiva en el hígado. Cada una de ellas realizaba una función distinta. El alma apetitiva controlaba los

impulsos naturales del cuerpo, como el hambre y la sed, los instintos de supervivencia y los placeres corporales, y estaba animada por espíritus naturales. El alma espiritual albergaba las emociones y las pasiones, y estaba animada por un espíritu vital que de alguna manera se formaba en el corazón a partir de la sangre y el aire que llegaba a través de los pulmones. El alma racional controlaba la cognición, como las percepciones, la memoria, la toma de decisiones, el pensamiento y la acción voluntaria. Galeno no veía distinción entre lo mental y lo físico. Puede decirse que empezó a sentar las bases de una serie de ideas tan modernas como el consciente frente al subconsciente, el yo y el ello, lo racional y lo intuitivo. Los detalles difieren, pero las ideas subyacentes empezaron a surgir ya en el año 200 de nuestra era.

Galeno hizo sus pinitos en el mecanicismo. Imaginó un espíritu vital, una fuerza vivificante que entraba en el cuerpo y se purificaba en la *rete mirabile*. Luego ese espíritu purificado fluía hacia los ventrículos del cerebro, donde se convertía en un espíritu animal y posibilitaba la cognición del alma racional. Aunque Galeno acertó en el órgano responsable de las funciones cognitivas, en realidad no entendió su funcionamiento, ya que localizó toda la actividad de procesamiento en los ventrículos vacíos, lo cual es como decir que la mejor parte de una rosquilla es el agujero.

Aun así, una de las mayores aportaciones de Galeno al futuro de la medicina fue la idea de que diferentes órganos realizan distintas funciones. Empezar a diferenciar los órganos del cuerpo en varias máquinas que realizan funciones diversas fue una idea extraordinaria. Hoy, uno de los objetivos de la neurociencia moderna es descubrir qué funciones desempeñan las diversas partes del cerebro. Un siglo tras otro, la neurociencia se va haciendo cada vez más específica con respecto a qué sistemas cerebrales concretos contribuyen al conjunto de nuestra vida mental. Al más puro estilo reduccionista, Galeno no diferenciaba entre lo físico y lo mental, pero a la vez se aferraba a la idea de un alma inmortal. Una y otra vez, como veremos, los brillantes precursores de la neurociencia moderna abandonaron su desapasionada capacidad de

razonamiento para, cual *deus ex machina*, introducir un espectro al final de su análisis.

Durante toda su vida, Galeno creyó firmemente en la observación y la experimentación personal por encima del saber establecido, pero no practicó por completo lo que predicaba. Su epistemología estaba arraigada en su formación filosófica, que incluía las enseñanzas de Platón, Aristóteles y los estoicos, de modo que mezcló y combinó parte de estas enseñanzas con sus observaciones para crear una teoría general de la medicina. Pese a ello, es muy probable que se hubiera sentido totalmente consternado de haber sabido la influencia que iba a ejercer en la medicina durante los mil trescientos años siguientes. Durante más de un milenio, los descubrimientos de Galeno se considerarían la biblia de la medicina, al tiempo que la nueva Iglesia cristiana llegaría a considerar como doctrinas algunas de sus ideas. Así, mientras que en el Antiguo Testamento el alma moría con el cuerpo, tal como afirmaba Aristóteles, los nuevos cristianos adoptaron una visión del alma completamente distinta: la concibieron inmortal, trascendente de la vida del cuerpo, como sugerían Platón y Sócrates. Aunque Galeno creía que no había distinción entre lo mental y lo físico, a los cristianos les gustó su idea de que el alma residía en los etéreos ventrículos cerebrales, lejos del alcance del cuerpo lujurioso y pecador. De modo que esta pasó a ser la doctrina de la Iglesia en lo relativo a la ubicación corporal de esa alma ahora inmortal e inmaterial. La sensación residía en el ventrículo frontal, el entendimiento ocupaba el ventrículo medio y la memoria cubría la retaguardia.

Desde los primeros griegos y durante todo el período de influencia de Galeno —en total, diecisiete siglos de pensamiento humano—, a la hora de concebir la naturaleza de la existencia humana nos hallamos sumidos en un enorme mar de confusión. Casi siempre se hablaba de almas, no de mentes, y, por supuesto, aún menos de conciencia. Platón y Sócrates apostaban por un alma inmortal tripartita, en parte racional, en parte espiritual y en parte apetitiva. Aristóteles también razonó que teníamos almas, pero afirmó que no

eran inmortales. Los primeros estudiosos del cerebro, y de la anatomía en general, volvieron a afirmar que eran inmortales, pero que no había diferencia entre lo mental y lo físico. Las ideas se resisten a desaparecer, aun a la luz de una ciencia emergente. Como veremos, esas primitivas ideas todavía siguen vigentes hoy en día.

Se sientan las bases para Descartes y la idea del dualismo mente/cuerpo

Habría que llegar al siglo XVI para que la anatomía de Galeno se viera cuestionada por un joven anatomista, Andrés Vesalio, que trabajaba en la Universidad de Padua. Vesalio se quedó perplejo cuando comparó sus propias disecciones humanas con los dibujos de Galeno. Afortunadamente para él —y para la ciencia moderna—, no tuvo que hacer frente a ningún tabú relativo a la disección de cuerpos humanos, y el juez local no tuvo el menor reparo en enviarle los cadáveres de los criminales condenados a muerte. Vesalio se dio cuenta de que Galeno no solo no había diseccionado nunca a un humano, sino que además una gran parte de su anatomía era sencillamente errónea. El joven anatomista no disponía precisamente de las mejores herramientas a la hora de diseccionar el cerebro. Lo serraba en rodajas de arriba abajo, estropeando las secciones inferiores conforme avanzaba (algo parecido a cortar una bola de *mozzarella* de búfala con un cuchillo romo). Pero una cosa quedó perfectamente clara: no había *rete mirabile*. Un aspecto de la ciencia cuya enorme importancia hemos descubierto con el transcurso de los siglos es el de comprobar y verificar a conciencia cualquier afirmación previa.

Unos años antes, otro anatomista llamado Niccolò Massa, de la Universidad de Bolonia, había descubierto que los ventrículos no estaban llenos de espíritus etéreos, sino de líquido. Vesalio, a su vez, descubrió que tampoco eran las esferas perfectas con bóvedas carnosas que describiera Galeno. Había tantos errores en las descripciones de Galeno que Vesalio tuvo

que reescribir (o redibujar) el libro, por así decirlo. Con la ayuda de varios aprendices del taller de Tiziano en Venecia, en 1543 se publicó *De humani corporis fabrica libri septem* («Sobre la estructura del cuerpo humano en siete libros»), donde aparecían esqueletos (con o sin sus músculos o sistemas circulatorios) paseando con bastones por la campiña italiana, recostados con aire despreocupado sobre columnas o troncos de árboles, o incluso observando atentamente libros apoyados en atriles. La obra tuvo un gran éxito, especialmente entre los estudiantes.

Tras haber despojado a tantos cadáveres de su piel, Vesalio pretendía salvar la suya. La estructura que supuestamente purificaba los espíritus vitales y los convertía en espíritus animales simplemente no estaba ahí. Y lo que resultaba aún más inquietante: los ventrículos que presuntamente albergaban el alma no estaban llenos de aire, ni se asemejaban a las descripciones que la Iglesia daba de ellos. Vesalio no cuestionaba su fe ni su alma inmortal, pero sabía que los padres de la Iglesia lo harían si desafiaba su doctrina, una jugada arriesgada en la época de la Inquisición. Él pensaba que quizá era en el cerebro en su conjunto, y no exclusivamente en los ventrículos, donde residía el alma (sensación, entendimiento y memoria). Fuera como fuese, usó la cabeza y guardó silencio.

A finales del siglo XVI los científicos pisaban el acelerador proporcionando nuevas observaciones. En Padua, Galileo no solo cuestionaba la noción aristotélica (y bíblica) de un universo centrado en la Tierra, sino que además utilizaba fórmulas matemáticas, mediciones y experimentos para demostrar que Aristóteles se equivocaba. El resultado fue que Galileo declaró que las leyes de la naturaleza —es decir, las leyes que gobiernan el mundo físico— eran matemáticas, es decir, mecanicistas. Acusado de pretender reinterpretar la Biblia, fue juzgado por la Inquisición romana, se le impuso silencio con respecto a su opinión sobre el Sol y fue sometido a arresto domiciliario.

Sin embargo, en París estaban surgiendo ideas nuevas. Marin Mersenne, un colega matemático además de teólogo, filósofo, teórico de la música y monje,

apoyaba a Galileo. Vivía en el convento de L'Annonciade, donde organizaba frecuentes debates en su celda con destacados pensadores y científicos de toda Europa, a la par que mantenía una extensa y profunda correspondencia con otros. Mersenne había decidido que, si la Iglesia había de sobrevivir a la embestida de la nueva ciencia y las reclamaciones de los herejes, tenía que aceptar y asimilar la idea de que el universo era mecanicista. Dios podía gobernar un universo que siguiera las leyes naturales que él mismo había creado tan fácilmente como uno centrado en el ser humano. De hecho, pensándolo bien, ¿por qué él, en su omnisciencia, no iba a crear un universo capaz de funcionar automáticamente sin necesidad de mantenimiento?

Entre los asistentes a aquellas sesiones se contaba otro filósofo-matemático-científico-sacerdote francés llamado Pierre Gassendi. Este suscribía la noción de que el mundo estaba compuesto de átomos, una teoría propuesta por primera vez en la cultura occidental en el siglo v a. C. por Leucipo y Demócrito. Los átomos se describían como indestructibles, inmutables y circundados por un vacío. Las diferentes clases de átomos tenían su propio tamaño y forma específicos, y todos se hallaban en constante movimiento. Los átomos podían unirse, y Gassendi denominó «molécula» a la estructura resultante, con distintas formas y diferentes propiedades intrínsecas. Todas las sustancias macroscópicas del mundo estaban compuestas de diversos átomos. Gassendi no consideró que aquella idea resultara herética en absoluto: Dios había creado los átomos al igual que todo lo demás.

En cambio, se equivocó al postular la existencia de dos clases de almas. Una de ellas estaba hecha de átomos, conectada al sistema nervioso y al cerebro, y era capaz de percibir, sentir placer y dolor y tomar decisiones. Sin embargo, había algo de lo que Gassendi estaba seguro: ningún átomo, ni ninguna combinación de ellos, podía reflexionar sobre sí mismo o percibir nada más allá de lo que le proporcionaba la sensación. Por lo tanto, concluyó que los humanos debían de tener otra alma, un alma racional que era inmaterial. Dicha alma, sin embargo, no era independiente: él creía que

durante la vida se fusionaba con el cuerpo y dependía de este para obtener información del mundo exterior. Tras la muerte, no obstante, el alma se revelaba inmortal y se escabullía.

Es aquí donde entra en juego el joven René Descartes, filósofo, matemático y racionalista, que también frecuentaba las sesiones de Mersenne y suscribía la idea de que el mundo físico estaba compuesto de partículas y funcionaba como una máquina. Dado a vestir con extravagancia y aficionado al tafetán, las plumas y las espadas, le gustaba pavonearse por París, que en aquella época contaba con su propia y visionaria versión de la atracción de Eurodisney «Qué pequeño es el mundo» en los Reales Jardines de Francia. Esta estaba integrada por una serie de autómatas hidráulicos que se movían, emitían sonidos y tocaban instrumentos musicales. Los autómatas se activaban mediante un ingenioso mecanismo que respondía a la presión ejercida sobre los senderos de losas de los jardines al caminar sobre ellos. En realidad los autómatas — más conocidos hoy en día como robots— eran bastante comunes por entonces y, sin duda, la mayoría de los visitantes de los jardines se sentían hechizados por ellos.

Descartes, sin embargo, era un filósofo, alguien para quien un paseo por el parque nunca era solo un paseo por el parque (de ahí el tafetán y las plumas). Sabía que aquellos robots de apariencia humana eran máquinas gobernadas por fuerzas externas inanimadas y, sin embargo, parecían realizar movimientos voluntarios racionales. Se le ocurrió pensar que ciertos aspectos de nuestro cuerpo eran muy similares. Nuestros reflejos son solo eso: un estímulo externo del entorno hace que suceda algo en el sistema nervioso que se traduce en una respuesta motora preprogramada. No es necesaria la intervención de nadie que dirija la acción. No hace falta ninguna alma. Descartes consideró asimismo que una respuesta refleja no tiene por qué ser únicamente una respuesta motora: podría ser también emocional o cognitiva, como un recuerdo. Una vez emprendido el camino por este particular sendero del pensamiento, las posibilidades teóricas de comportamientos generados por algún tipo de

reacción refleja a un estímulo externo eran ilimitadas. Pero a la vez eran deterministas: el estímulo x producirá siempre la reacción y . Descartes aprobaba esa idea para las máquinas y los animales, pero ¿se aplicaba también a los humanos? ¿No había libre albedrío? ¿Ni decisiones voluntarias? ¿Ni responsabilidad personal por nuestros actos? ¿Ni moral ni pecados? ¿Éramos nosotros mismos máquinas? Eso era demasiado.

Retrocediendo ante el vacío de tal desesperanza existencial, Descartes empezó a desarrollar una idea que cambiaría la historia. Pero el daño al estudio de la biología ya estaba hecho. El distinguido biólogo teórico Robert Rosen señala que, si bien nadie puede explicar qué es un organismo viviente, resulta fácil decir cómo es. Rosen sostiene que Descartes lo entendió al revés: «Procedió a invertir la relación entre aquellos autómatas y los organismos que simulaban. Lo que había observado era simplemente que los autómatas, en las condiciones adecuadas, a veces pueden parecer vivos. Y lo que concluyó fue, en cambio, que *la propia vida era como los autómatas* [la cursiva es de Rosen]. Así nació la metáfora de la máquina; probablemente, la mayor fuerza conceptual en biología, aun hoy». ³ Y nació también el mundo absolutamente determinista que esta implica.

Es cierto que nuestro cuerpo mueve involuntariamente la pierna hacia arriba cuando le dan un golpecito en la rodilla, pero también podemos realizar el mismo movimiento de forma voluntaria. Se trata de dos casos muy distintos: en uno el cuerpo reacciona a un estímulo externo; en el otro, según Descartes, es la mente la que inicia el movimiento. Mientras que es posible describir el primero mecánicamente utilizando las leyes de la física como una cadena de acontecimientos que pueden remontarse a la creación, el segundo, en opinión del filósofo, era una cadena causal de dos eslabones: quieres hacerlo y *voilà!*, ocurre. ¿Por qué lo has hecho? Porque querías: no hay aquí nada físico que estudiar. Solo un deseo. Lo que Aristóteles denominara «causa final».

Descartes rechazó la idea de que los acontecimientos voluntarios fueran un reflejo o mecanismo físico que pudiera describirse científicamente.

Finalmente llegó a la conclusión de que, aunque el cuerpo estaba gobernado por leyes físicas, lo que causaba o impulsaba la acción humana era un agente autónomo rector, el alma racional, que no estaba hecho de materia, es decir, que no era físico, ni mecanicista, ni estaba constreñido por ninguna ley natural; era algo salido de la nada. Esta alma estaba dotada de conciencia, de libre albedrío y de pensamiento abstracto, y era capaz de albergar dudas y moral. Esto es lo que se conoce como «dualismo mente/cuerpo»: la idea de que el cuerpo consiste en maquinaria física y la mente en maquinaria cognitiva no física (inmaterial).

Descartes era un matemático y científico con todas las de la ley, y, como tal, quería entender racionalmente la verdadera naturaleza del ser. Dado que sus enfoques matemáticos racionales funcionaban bien para el mundo físico (había desarrollado la geometría analítica y descubierto la ley de la refracción, entre otras cosas), intentó abordar la verdadera naturaleza del hombre utilizando el mismo método racional. Para ello, primero tenía que descartar todo aquello de lo que podía dudar a fin de obtener una certeza, una base sobre la que construir sus argumentos. Resultó que lo que encontró fue una manera de dudar de casi todo, incluso de que su madre era su madre, de que el sol saldría al día siguiente o de que la noche anterior había dormido en su cama en París en lugar de andar pavoneándose por Roma. Incluso podía dudar de que tuviera un cuerpo. Al fin y al cabo, la creencia de que tenemos un cuerpo se basa en percepciones sensoriales, las cuales a veces resultan erróneas. Y si son erróneas una vez, bueno, entonces también podrían serlo siempre. Sin embargo, había algo que Descartes sabía con absoluta certeza, algo de lo que no podía dudar: sabía a ciencia cierta que existía. En su propio proceso de duda estaba afirmando que era un ente pensante. Por lo tanto, *cogito ergo sum*: «pienso, luego existo».

Ahora que Descartes consideraba que tenía unos cimientos sólidos sobre los que edificar, se propuso deducir de una vez por todas la verdadera naturaleza del ser, y hacerlo paso a paso, científicamente. Entonces hizo la

reflexión de que, dado que podía dudar de que tuviera un cuerpo, también podía dudar de su propia existencia física. Siguiendo ese fino hilo de razonamiento, concluyó: «Conocí por ello que yo era una sustancia cuya esencia y naturaleza toda es pensar, y que no necesita, para ser, de lugar alguno, ni depende de cosa alguna material; de suerte que este “yo”, es decir, el alma, por la cual yo soy lo que soy, es enteramente distinta del cuerpo y hasta más fácil de conocer que este; y aunque el cuerpo no fuese, el alma no dejaría de ser cuanto es». ⁴ Su pensamiento siguió por ese tortuoso camino, extrayendo conclusiones de argumentos en los que, desde nuestra perspectiva actual, resulta fácil encontrar defectos. Por ejemplo, no es difícil ver que, solo por el mero hecho de que uno pueda dudar de que existe como ente físico, ello no implica necesariamente que esté en lo cierto y que realmente no sea un ente físico, ni que el cuerpo no sea esencial para los pensamientos. No obstante, tal era la inestable base del primero de los argumentos de Descartes que sustentarían el dualismo mente/cuerpo.

Pero los argumentos de Descartes no gozaban de las ventajas del conocimiento actual; sus conclusiones e ideas modelarían el pensamiento intelectual hasta la época moderna, y su dualismo mente/cuerpo, su afirmación de la separación entre la mente, por una parte, y el cuerpo y el cerebro, por otra, ha ejercido un dominio absoluto sobre la filosofía de los últimos trescientos cincuenta años. En cambio, sus coetáneos tuvieron problemas para aceptar sus conclusiones. Muchos de sus partidarios, incluida la princesa Isabel de Bohemia (cuya correspondencia con Descartes fue muy extensa), se preguntaron cómo esa mente inmaterial interactuaba con el cuerpo material. Descartes le confesó a Isabel que no tenía una buena respuesta (puede que le hubiera reconfortado saber que la cuestión es objeto de debate todavía hoy). ⁵ De todos modos lo intentó: buscó en el cerebro y encontró lo que él creyó que era el lugar donde residía la interacción entre este y la mente: la glándula pineal. De modo que escribió a Isabel: «Mi punto de vista es que esta glándula es la sede principal del alma y el lugar en el que se forman todos

nuestros pensamientos. La razón por la que lo creo es que no puedo encontrar ninguna parte del cerebro que no esté duplicada, salvo esta». ⁶ Aquí uno no puede menos que preguntarse si no estaba agarrándose a un clavo ardiendo; al fin y al cabo, su investigación consistió en examinar cerebros de terneros, de los que ya había afirmado que carecían de alma inmaterial, y los erróneos dibujos de Galeno.

Mientras trabajaba en todo esto, Descartes deslizó la palabra *consciente* una sola vez en sus escritos, concretamente en el párrafo 32 de la tercera de sus *Meditaciones*, introduciendo así el término en la filosofía. Como todos los hombres cultos de su época, escribía en latín, de modo que en realidad utilizó la palabra latina, *consciuis*. Posteriormente la traducción al francés no sería tan estricta con sus interpretaciones o con el uso de la palabra, empleándola también cuando el propio Descartes utilizaba los verbos «pensar», «saber» o «conocer». Sea como fuere, de inmediato se plantearon objeciones al uso del término. Puede que el propio Descartes se arrepintiera de haber llegado a usarlo, ya que luego se dedicó a dar mil vueltas en torno a su significado, vacilando con respecto a si la conciencia era reflexiva —es decir, un pensamiento sobre un pensamiento— o era simplemente el pensar en general. En cualquier caso, Descartes utilizó el término para referirse al conocimiento que tenemos de lo que está pasando en nuestra mente; un conocimiento que para él era a la vez indubitable e infalible, una conclusión a la que llegó por razonamiento lógico. Así, por ejemplo, si yo pienso que tengo el mejor viñedo del mundo, no tengo la menor duda de que eso es lo que estoy pensando: es un hecho indubitable. Además, no puedo equivocarme al afirmar que lo que estoy pensando es precisamente eso: es, pues, infalible. Como Descartes sabía a ciencia cierta lo que estaba pensando, eso significaba que conocía su mente mejor que su cuerpo: su conciencia no podía engañarlo.

Descartes y los franceses dieron origen a una industria filosófica que desde entonces se ha esforzado en dar sentido a una idea de conciencia que de entrada nunca se definió claramente. En última instancia no era un caso muy

distinto de la famosa opinión sobre la pornografía vertida por el juez del Tribunal Supremo estadounidense Potter Stewart: «No intentaré [...] seguir definiéndola [...] y quizá nunca lograría hacerlo de una forma inteligible. Pero la reconozco cuando la veo».

Dejamos la Francia del siglo XVII equipada con un universo mecanicista y dos descripciones distintas de la mente. Antes de Descartes, la noción de un alma, ya fuera material o inmaterial, dominaba el pensamiento humano. Era como si la presencia consciente que los humanos sentimos y experimentamos hiciera casi imposible pensar que nuestra «alma» era un trozo de carne. Comprensiblemente, resulta difícil, y hasta francamente enojoso, pensar que después de toda una vida de esfuerzo la fiesta se acaba con la muerte. Aristóteles trató de ponernos en el camino correcto en estas lides, dejando claro que, con la muerte, el alma también moría. Sin embargo, aun después de dos mil años de acumulación de conocimiento humano, la mayoría de las personas siguen sin suscribir la sencilla realidad de que son nuestros cuerpos (y nuestros cerebros) los que generan lo que somos, en toda nuestra complejidad biológica y cultural.

En esta ruta hacia el presente, Descartes separó audazmente el alma inmortal (y, con ella, la mente) del universo y el cuerpo mecanicistas. Con la mente y la carne consideradas como entidades distintas, la primera pasó a convertirse en el enigma central; se la consideró inmaterial, indubitable, infalible e inmutable. Al dotar a la mente de un estatus sobrenatural, Descartes la eliminó como objeto de estudio científico. Nunca fue capaz de explicar cómo esa mente inmaterial interactuaba con el cuerpo material, pero su teoría paralizaría profundamente el pensamiento sobre la realidad física de la mente durante más de doscientos años. Muchos de sus más brillantes contemporáneos, como Pierre Gassendi, coincidieron en que había un alma racional inmaterial porque estaban seguros de que ningún átomo ni combinación de ellos podía reflexionar sobre sí mismo o percibir nada más

allá de lo que le proporcionaba la sensación. Por extrañas e inútiles que estas ideas del siglo XVII fueran en su momento, la noción de que existen estados mentales sigue vivita y coleando en la ciencia del XXI. Solo que, en lugar de una mente inmaterial flotando en cada uno de nosotros, la ciencia moderna ha trasladado esa mente al cerebro y la ha convertido en algo físico. Pero la pregunta que sigue en el aire es: ¿cómo diablos funciona?

2

Los albores del pensamiento empírico en filosofía

—No creo...

—Entonces no deberías hablar —dijo el Sombreroero.

LEWIS CARROLL, *Alicia en el País de las Maravillas*

En la orilla del Canal de la Mancha opuesta a Descartes y sus compañeros parisinos, los ingleses también se devanaban los sesos en torno al significado de la vida, el alma y la mente. La palabra *conciencia* cuajó entre los filósofos británicos, y, cincuenta años después de que Descartes la acuñara en sus *Meditaciones*, John Locke amplió su significado, como hizo también el escocés David Hume. Pero los filósofos no estuvieron solos en esta empresa. El mundo de la medicina, con su inherente interés en el cuerpo y la anatomía, también empezó a explorar los temas de la mente y el cerebro. Thomas Willis y William Petty trabajaban arduamente en Oxford, y sus hallazgos estaban a punto de influir en el efervescente debate sobre el dualismo cerebro/mente. En cierta medida se repitió la misma historia de siempre. Los científicos eran hijos de la religión. Su nuevo conocimiento científico del mundo chocaba con las sinceras creencias religiosas de su infancia. Experimentaban lo que hoy se conoce como «disonancia cognitiva», la incomodidad mental que uno siente cuando alberga a la vez dos o más creencias, ideas o valores contradictorios. Como resultado, para reducir esa incomodidad, la gente intenta explicar o justificar el conflicto, o bien, en su lugar, cambia directamente sus creencias. En la época de la que hablamos casi todo el mundo sentía un acuciante deseo

de que la creencia en Dios no fuera una víctima de los descubrimientos de su joven ciencia. De modo que, para reconciliar su pensamiento sobre la mente, de la que sabían muy poco, con su pensamiento acerca del cuerpo, sobre el que estaban aprendiendo cada vez más, aquellos científicos empezaron a hacer sugerencias bastante absurdas en lo relativo a la interacción entre ambos. De hecho, en un primer momento los neurocientíficos de la época se sintieron tan desconcertados como los filósofos por su propia percepción de la conciencia y su nuevo compromiso con el pensamiento objetivo.

Al manantial de ideas que brotaban en Francia e Inglaterra se sumó una avalancha de trabajo de los alemanes. Desde Leibniz hasta Kant, el continente bullía de ideas sobre la naturaleza de la mente. El mero hecho de ver las ideas formarse, transformarse y cambiar ya es una maravilla en sí. Descartes, con toda su genialidad y confianza en sí mismo, había arrojado el guante al proponer que la mente no está hecha de la misma sustancia que el cerebro. Este acto intelectual resultaría ser un reto para las mentes sofisticadas y perspicaces de los doscientos años siguientes. En muchos aspectos, aquel largo debate sería una auténtica batalla campal, y su importancia sería deslumbrante.

La tabula rasa, la experiencia humana y los comienzos de la neurociencia

A mediados del siglo XVII, Inglaterra se vio envuelta en una despiadada guerra civil en torno a la religión y el poder del monarca. Thomas Hobbes, que era monárquico, además de un erudito paradigmático, había regresado a París desde Londres, frustrado por la mala acogida que había tenido un librito que había escrito sobre la política de la época. En París encontró trabajo como tutor del exiliado príncipe Carlos (el futuro Carlos II), y no tardó en convertirse en uno de los habituales de la tertulia de Mersenne. Desde un primer momento, Hobbes, que tenía formación en física, pareció desentenderse

del alma inmaterial. Rechazaba de plano el concepto de alma de Descartes, que consideraba una ilusión. Él creía que lo que posibilitaba la razón no era una misteriosa «sinsustancia», sino simplemente la capacidad del cuerpo de mantener el orden en el cerebro. Pensaba como un ingeniero: constrúyelo, hazlo funcionar, y eso es todo; no hay fantasmas en el sistema.

Hobbes no daba abasto, ya que, además de hacer de tutor del príncipe, estaba escribiendo dos libros: uno sobre la visión y otro sobre el cuerpo y su maquinaria. Necesitaba un ayudante, y contrató a un joven e inteligente estudiante de medicina británico llamado William Petty. Por alguna razón, Hobbes tenía la idea preconcebida de que los sentidos generaban una presión que era la causante de los latidos del corazón. El joven Petty le ayudó a estudiar los libros de Vesalio, pero Hobbes no encontró evidencia alguna que sustentara su teoría. Pese a ello, y en consonancia con la naturaleza básica de muchos científicos, no desfalleció.

Hobbes asistió a disecciones en compañía de Petty, esperando ver brotar los nervios del corazón como si fueran las púas de un erizo de mar, y, desde allí, extenderse en todas direcciones. Pero no había tales. Cuando finalmente comprendió que estaba equivocado, Hobbes abandonó su hipótesis. En la ciencia, como en la vida, nuestro entorno social nos brinda la oportunidad de intercambiar ideas con los demás, un intercambio que es de naturaleza bidireccional. El cambio de opinión de Hobbes impresionó tanto a Petty que este decidió adoptar su mismo método, explorar extensamente una cuestión y a la vez mostrarse flexible ante los resultados: si sus suposiciones no coincidían con las observaciones, cambiaría de opinión. A su regreso a Inglaterra, Petty se llevó consigo un obsequio material de Hobbes bajo el brazo, un microscopio, pero también un legado conceptual en la cabeza: la convicción de que el cuerpo era un conjunto de piezas que funcionaba como una máquina. Aun así, el mayor regalo de Hobbes a Petty fue el de instarle a comprender el valor de responder a una pregunta mediante la observación y la experimentación, en lugar de tergiversar las observaciones para hacerlas

encajar con las propias teorías. Pero, puede creerme, eso resulta más fácil de decir que de hacer: a nadie le gusta admitir que se ha equivocado.

Petty se convirtió en un anatomista excepcional. No mucho después de su regreso a Inglaterra se instaló en Oxford, donde, como Vesalio antes que él, pasó a disponer de un constante flujo de cadáveres procedentes de la horca. Allí se le unió otro joven médico, Thomas Willis, monárquico y ferviente anglicano. Debido a estas ideas, que a escala local no resultaban demasiado populares, la formación de Willis había sido bastante irregular. Petty lo compensó con creces, y durante los cinco años siguientes convirtió a Willis en otro extraordinario anatomista, inclinado, como él, a dar prioridad al aprendizaje basado en la observación y la experimentación. El jovencísimo campo de la neurociencia estaba dando sus primeros pasos en Gran Bretaña cuando estos dos científicos tomaron las riendas. Al poco tiempo nadie podría ignorar el carácter esencial del cerebro a la hora de pensar en los estados mentales, la conciencia y todavía, para algunos, el alma.

Un pequeño paso para la ciencia

Hay muchas cosas que contribuyen a labrarse una gran reputación científica, especialmente cuando un ámbito de estudio es joven y aún no se ha puesto a prueba. Petty y Willis tuvieron suerte cuando, un año después de que este último iniciara su trabajo, llegó un ataúd a su laboratorio con la última víctima de la horca, Anne Greene, una mujer que había sido violada y luego condenada a muerte por matar a su bebé recién nacido. Había permanecido colgada del cuello durante media hora y, como era habitual en aquella época, sus amigos se habían agarrado a su cuerpo mientras colgaba de la soga, utilizando su peso para acelerar su muerte. Al día siguiente, a la hora de hacer la autopsia, el laboratorio de Petty no tardó en llenarse de gente.

Con Petty al timón, las disecciones se habían convertido en algo parecido a un espectáculo. Pero, antes de que entrara en la habitación, alguien levantó la

tapa del ataúd y entonces se oyó una especie de gorgoteo procedente del interior al más puro estilo de Edgar Allan Poe. Cuando Petty y Willis entraron por la puerta uno de los espectadores estaba dando fuertes pisotones en el pecho de Greene. Trataron frenéticamente de revivirla por diversos medios y tuvieron éxito, hasta el punto de que a la mañana siguiente Greene estaba pidiendo cerveza. Los magistrados querían colgarla de nuevo, pero los dos médicos los convencieron de que en realidad había tenido un aborto espontáneo (solo había estado embarazada cuatro meses) y el bebé ya estaba muerto al nacer. Fue absuelta y más tarde llegaría a tener tres hijos. Aquel espectacular acontecimiento trajo fama y fortuna a Petty y Willis, y les garantizó una envidiable vida dedicada a la investigación sin necesidad de tener que ir mendigando apoyo financiero.

Petty y Willis siguieron colaborando durante otros cuatro años. Bajo la tutela de Petty, Willis empezó a realizar autopsias a sus pacientes cuando morían, a entender mejor el cuerpo y cómo este se veía afectado por las diversas enfermedades, y quizá también a descubrir qué las causaba. Más tarde Petty partió en busca de mejores perspectivas y viajó a Irlanda como médico con el ejército de Oliver Cromwell; con el tiempo llegaría a ser un conocido economista, parlamentario y miembro fundador de la Royal Society de Londres. Willis tomó el relevo y empezó a interesarse especialmente por el cerebro, desarrollando técnicas de disección que le permitieron conocer la anatomía de este con mayor claridad que sus predecesores. En colaboración con Christopher Wren —quien, entre otros logros (fue astrónomo, cirujano y arquitecto), fue pionero en el arte de inyectar tintes en las venas—, Willis perfiló el sistema vascular del cerebro inyectando tinta y azafrán en la arteria carótida de un perro. Fue el primero en comprender la función de la estructura vascular de la base del cerebro, que, en su honor, pasaría a llamarse «polígono de Willis».

Juntos, Willis y Wren elaboraron los dibujos más precisos del cerebro humano realizados hasta entonces, que luego publicaron en un libro *The*

Anatomy of the Brain and Nerves («La anatomía del cerebro y los nervios»). El libro se agotó muy pronto, y en el plazo de un año se hicieron cuatro ediciones. Habrían de pasar más de doscientos años para que se llegara a superar la calidad de sus dibujos anatómicos.

Pese a todos sus conocimientos de anatomía, Willis seguía aferrado a la idea de que había unos espíritus vitales y sensibles que mantenían vivo el cuerpo, una idea del pasado que no tenía visos de desaparecer. Pero Petty lo había formado bien, y a la larga cambió de opinión cuando sus estudiantes lo convencieron, a través de una serie de inteligentes experimentos, de que no había ningún espíritu en juego. La sangre recogía algo del aire y lo llevaba a los músculos: esa era la fuerza impulsora del cuerpo. No llegaron a detectar que se trataba de un elemento químico, pero estuvieron muy cerca.

Tras realizar numerosas disecciones de animales, Willis vio que había estrechas similitudes entre el cerebro de estos y el de los humanos. A partir de sus observaciones llegó a la conclusión de que el alma humana y la de los animales eran muy parecidas, y diferían en aspectos que solo podían observarse en sus cuerpos. Así, por ejemplo, los animales con un bulbo olfatorio más grande tenían más desarrollado el olfato. Willis vio que los humanos tenían una corteza cerebral mucho mayor que la de otros animales, y concluyó que esta debía de ser la sede de la memoria, dada nuestra mayor capacidad de recordar. Aunque este pueda parecer un argumento tosco y simplista, la conclusión de Willis no era muy distinta de algunas de las ideas más prometedoras que circulan en el ámbito de la neurociencia moderna. De hecho, el Premio Kavli de Neurociencia 2016 fue otorgado a Michael Merzenich, el científico que ha demostrado que las áreas del cerebro asociadas con determinadas actividades aumentan de tamaño con el uso.

No obstante, las disecciones de animales habían planteado un gran problema a Willis. Aunque los humanos son capaces de pensar de maneras enormemente distintas de otros animales, la organización de sus cerebros parecía ser muy similar. Como no podía encontrar ninguna sustancia cerebral

material que explicara esa diferencia, su conclusión lógica fue que debía de haber otra cosa que otorgaba a los humanos aquella capacidad: un alma racional. De modo que volvíamos a las andadas. Como no lograba ubicar físicamente el pensamiento racional en el cuerpo, aceptó la visión de Gassendi de que se trataba de algo inmaterial, aunque estaba localizado en el cerebro, tal como afirmaba Descartes. Willis dedujo que los nervios captaban sensaciones del mundo exterior y los espíritus animados los llevaban al cerebro. Dentro del cerebro, los espíritus seguían trayectorias que los internaban profundamente hasta un punto de encuentro general: el enorme paquete de nervios que conecta los dos hemisferios cerebrales, el denominado cuerpo calloso. Así, una vez más, una mente brillante volvía a malinterpretar la cuestión clave. Es como si un científico moderno examinara el interior de un ordenador, no viera nada especial, y concluyera que debía de haber un espíritu inmaterial gravitando en torno a la placa madre que era el que lo hacía funcionar.

El alma reina, pero no gobierna

Willis, como buen monárquico, veía esta alma racional como el «rey» del cuerpo. Como el jefe de cualquier gran organización, el rey solo disponía de la información que le llegaba, y no tenía conocimiento directo del mundo exterior. Y al igual que en cualquier estructura de este tipo, dicha información podía ser defectuosa o no estar disponible. Debido a que el propio cerebro en sí es un órgano físico, él o sus partes podrían enfermar y no proporcionar la información adecuada, afectando con ello a la línea de abastecimiento de información al alma racional. De modo que, cuando el cerebro está enfermo, existe la posibilidad de que el alma racional pueda verse afectada, a veces de forma permanente. Esta era, y sigue siendo, una idea potente. Como cabría esperar, Willis describió varias enfermedades mentales que había encontrado en sus pacientes para respaldar su teoría.

La figura de Willis es importante en nuestra búsqueda de la conciencia porque fue uno de los primeros en vincular lesiones cerebrales concretas a déficits conductuales específicos, y porque comprendió que diversas partes concretas del cerebro realizaban tareas distintas. En su libro *Two Discourses Concerning the Soul of Brutes* («Dos discursos sobre el alma de los brutos»), describió un cerebro que realizaba varias tareas de forma autónoma, no en una sola ubicación, sino distribuidas en toda su extensión; también describió canales de comunicación, aunque en aquella época —anterior al descubrimiento de la electricidad— imaginó que a través de ellos fluían espíritus en lugar de una señal eléctrica. En cualquier caso, fue quien puso en marcha los engranajes de lo que hoy es el campo de la neurociencia cognitiva, la moderna ciencia que ha asumido la tarea de intentar comprender la experiencia consciente humana.

La eclosión de John Locke

Se cree que tanto el trabajo empírico de Willis sobre anatomía como su teorización ejercieron una gran influencia en la formación de John Locke, por entonces aún un filósofo en ciernes. Locke también empezó como médico y estudió en Oxford con Willis. Entre las pocas clases a las que Locke consideraba que valía la pena asistir figuraban las clases de anatomía de Willis. Posteriormente, Locke trabaría amistad con otro médico, Thomas Sydenham, que también había sido compañero de clase de Willis. Sydenham había adquirido gran parte de sus conocimientos médicos mediante la experiencia de primera mano, el tipo de método que Locke llegaría a considerar fundamental.

Visitando a centenares de pacientes, Sydenham observó que cada enfermedad tenía su propio conjunto de síntomas sin importar quién fuera el paciente, un herrero de Sussex o el propio duque de York. Llegó a la conclusión de que las enfermedades podían diferenciarse unas de otras por su

lista de síntomas característica, y empezó a clasificarlas como si fueran plantas. Aquella era una idea revolucionaria, ya que hasta la fecha todavía seguían siendo populares los diagnósticos y tratamientos de Galeno, que eran mucho más subjetivos. Para Galeno, la enfermedad de cada persona estaba causada por un desequilibrio único de humores que requería un tratamiento a medida. Sydenham, en cambio, dio origen a lo que hoy consideraríamos los fundamentos de una medicina empírica. Para cada enfermedad probaba diferentes tratamientos en un grupo de pacientes, y luego evaluaba y ajustaba la medicación de acuerdo con su eficacia. Irónicamente, la visión de Galeno se halla más en consonancia con el reciente entusiasmo de la medicina contemporánea por los tratamientos personalizados, mientras que la opinión de Sydenham sintoniza más con la medicina algorítmica, que viene dictada por la adopción de prácticas y procedimientos estándar para todos los pacientes con los mismos síndromes concretos. En los capítulos 7 y 8 veremos que es habitual en ciencia encontrar diferentes enfoques, como estos, que se traducen en grandes batallas en torno a dos alternativas excluyentes, cuando en realidad existe como mínimo una opción alternativa. Estas situaciones se conocen como «falsos dilemas», un tipo de falacia informal. Quizá de forma hasta cierto punto inesperada para el lector, descubriremos que han sido los físicos quienes han tomado la delantera a la hora de demostrar que normalmente no basta ninguna de las respuestas por sí sola. Esta idea nos será muy útil cuando consideremos cómo las neuronas crean mentes.

Como cabría esperar de un futuro filósofo, Locke interpeló a Sydenham sobre sus métodos. ¿Realmente necesitaba conocer la causa primera de una enfermedad para tratarla? ¿Era posible siquiera saber cuál era? Mientras que Willis creía que las causas eran cognoscibles y perseguía ese objetivo en todas sus disecciones y experimentos, Locke y Sydenham no pensaban igual. Ellos llegaron a la conclusión de que las causas de la enfermedad estaban fuera del alcance del entendimiento humano, y posteriormente Locke se convencería de que el funcionamiento de la mente y la esencia de las cosas

eran igualmente insondables. Locke abordaría la filosofía del mismo modo que Sydenham abordaba la enfermedad: limitándose a hablar de nociones basadas en experiencias cotidianas. Dada su postura, no es de extrañar que al filósofo se le ocurriera la célebre idea de la *tabula rasa*, o «tabla rasa», según la cual la mente se forma únicamente a partir de la experiencia y la autorreflexión. Esta formulación constituye el fundamento de la teoría estándar del hombre que actualmente sustentan las ciencias sociales: la cultura se impone a la naturaleza.

Si bien Locke y Descartes terminarían ambos siendo dualistas, diferían en muchos detalles. Abordando la cuestión del alma desde una perspectiva psicológica, escribía Locke: «la conciencia es la percepción de lo que pasa en la propia mente de un hombre». ¹ Para él, dicha percepción, o consciencia de una percepción, se logra mediante un «sentido interno» que él denominaba «*reflexión*», porque las ideas que ofrece son solo aquellas que la mente consigue al reflexionar sobre sus propias operaciones dentro de sí misma». Locke incluso llega a afirmar la imposibilidad de que existan estados mentales inconscientes: «Es imposible que uno perciba sin percibir que lo hace».

Contrariamente a Descartes, Locke cortó la conexión entre el alma y la mente (entendiendo por mente aquello que piensa). Recuérdese que para Descartes alma y mente son una misma cosa: el pensamiento es el principal atributo de la mente, y una sustancia no puede existir sin su principal atributo; en consecuencia, la mente piensa constantemente, incluso cuando duerme, aunque esos pensamientos se olviden de inmediato. Locke también aceptaba que la mente despierta nunca deja de tener pensamientos, pero, basándose en su experiencia, rechazaba la idea de que el durmiente sin sueños los tuviera. Sin embargo, era obvio que el durmiente debía de seguir teniendo su alma; de no ser así, ¿qué ocurriría si moría mientras estaba dormido? De modo que, para Locke, la mente (con su propiedad de la conciencia) y el alma tenían que ser independientes. ¡Simple!

Descartes también limitaba el contenido de la conciencia a las operaciones

actuales de la mente. Locke no ponía tales límites. Para él, uno puede ser consciente de operaciones y acciones mentales pasadas. Locke concebía la conciencia como el pegamento que aglutina la propia historia dando lugar a la percepción de uno mismo, a la identidad personal. Creía que la conciencia nos permite reconocer nuestras experiencias pasadas como propias. Aunque coincidía con Descartes en que los humanos tienen libre albedrío, eludía la cuestión de cómo la materia podía producirlo introduciendo un Dios omnipotente en la ecuación y afirmando que este lo había hecho así.

Ahí estaban, pues, algunos de los hombres más inteligentes del mundo diseñando modelos que describieran cómo debían de funcionar el cuerpo, la mente y el alma; pero tenían que encajar lo que a ellos les parecían realidades indiscutibles en un marco que diferenciaba los humanos, los animales, los dioses, las mentes y la conciencia. Eso es lo que hacen aún hoy quienes diseñan modelos teóricos: crean un modelo y luego lo reajustan constantemente, con reajustes supuestamente basados en nueva información, hasta que dicho modelo parece explicar un problema. En este caso, no obstante, el modelo era un desastre.

A finales del siglo XVII, las ideas acerca de qué es la conciencia eran tan abundantes como confusas. Se iban acumulando datos que algún día proporcionarían marcos empíricos a los teóricos del futuro, pero todavía faltaba una estructura conceptual básica y exhaustiva capaz de explicar la conciencia. En suma, los filósofos discrepaban en torno a la propia idea, y algunos incluso pensaban que la concepción filosófica de conciencia era incoherente. La tarea de encarrilar la idea de conciencia en el camino recto hacia el futuro recaería en un precoz filósofo escocés llamado David Hume, quien a los dieciocho años de edad ya estaba hastiado de las «interminables disputas» de la filosofía. Hume consideraba la filosofía moral y natural de los antiguos «completamente hipotética, y basada más en la invención que en la experiencia». ²

El camino hacia las concepciones modernas

Hume pareció irrumpir en escena como un iconoclasta prefabricado, listo para poner fin a toda la palabrería del idealismo. Él creía que la idea de que la mente era de algún modo sobrenatural al cuerpo era una ilusión, y además una ilusión absolutamente tonta. Quería acabar con ella y estructurar una ciencia de la naturaleza real de la vida. Y eso fue exactamente lo que hizo, reencauzando así el pensamiento humano sobre la naturaleza de la mente para los siglos venideros.

Hume no tardó en darse cuenta de que las mismas falacias que tan extendidas estaban en el mundo antiguo —como partir de hipótesis basadas en la especulación y la invención, en lugar de la experiencia y la observación— también se hallaban presentes entre sus contemporáneos. Él creía que nuestro conocimiento de la realidad se basa en nuestra experiencia y, para bien o para mal, en los axiomas que elegimos. Los axiomas son enunciados que parecen tan evidentes o consolidados que se aceptan sin controversia o cuestionamiento, y simplemente se afirman sin tener pruebas que los sustenten. En pocas palabras, un axioma es una suposición u opinión básicamente indemostrable. El problema de basar el conocimiento en un axioma es que entonces resulta que lo que uno concluye acerca de la realidad depende de los axiomas que haya elegido. Como advierte el físico Robert Brown, de la Universidad Duke, «no hay nada más peligroso ni más potente en el proceso filosófico que la selección de los propios axiomas [...]. No hay nada más inútil que entablar un debate filosófico, religioso o social con otra persona cuyos axiomas difieren significativamente de los propios».³

De hecho, Hume llegó a la conclusión de que muchas de las preguntas que formulaban los filósofos eran «seudopreguntas», es decir, preguntas que no podían responderse con herramientas como la lógica, las matemáticas y la razón pura porque sus respuestas siempre se fundamentarían, en mayor o menor medida, en una creencia indemostrable, en un axioma. Él opinaba que los filósofos debían dejar de perder el tiempo —y hacérselo perder a los

demás— escribiendo copiosamente sobre seudopreguntas, prescindir de sus apriorismos y refrenar sus especulaciones, tal como hacían los científicos. Tenían que rechazar todo lo que no se basara en hechos u observaciones, y eso incluía eliminar cualquier recurso a lo sobrenatural. Hume apuntaba a Descartes y a otros como él que creían haber demostrado de manera concluyente la filosofía dualista mediante la razón, las matemáticas y la lógica. Hoy la postura de Hume es relativamente frecuente, en parte porque los filósofos académicos contemporáneos son contratados por modernas universidades dedicadas a la investigación y trabajan en un entorno científico experimental; aunque las ideas cartesianas siguen presentes, la mayoría de los filósofos y científicos ya no las toman en serio. Pero a principios del siglo XVIII el ataque de Hume a Descartes resultaba tan audaz como innovador.

El grandioso plan de Hume consistía en crear una «ciencia del hombre»; es decir, entender las leyes fundamentales que guiaban el funcionamiento de la mente en consonancia con lo que se sabía del mundo newtoniano y utilizando el método científico de Newton. Él creía que entender la naturaleza humana, incluidas sus capacidades y fragilidades, nos permitiría comprender mejor las actividades humanas en general; y también nos permitiría apreciar las posibilidades y los escollos de nuestras actividades intelectuales, incluyendo qué aspectos de nuestro pensamiento podrían limitar nuestros intentos de comprendernos a nosotros mismos. De hecho, consideraba que su ciencia del hombre debía ocupar un lugar prioritario por encima de las ciencias newtonianas, y escribió: «Hasta las matemáticas, la filosofía natural y la religión natural dependen en parte de la ciencia del hombre, pues se hallan bajo el conocimiento de los hombres y son juzgadas por sus poderes y facultades. Es imposible decir qué cambios y progresos podríamos hacer en estas ciencias si conociéramos totalmente la extensión y la fuerza del entendimiento humano». ⁴ Como suele decirse, Hume estaba en ello, aportando algo de rigurosa claridad al lío del dualismo mente/cuerpo. Algunos lo consideran el padre de la ciencia cognitiva.

Corría 1734, y a los veintitrés años de edad Hume asistió al *alma mater* de Descartes, el Colegio Jesuita de La Flèche, en Anjou, Francia. En su tiempo libre escribió el ya clásico *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects* («Tratado de la naturaleza humana: ensayo para introducir el método del razonamiento experimental en los asuntos morales»), que en 1748 reformuló, pulió, corrigió y clarificó en una nueva publicación que llevaba por título *An Enquiry Concerning Human Understanding* («Investigación sobre el entendimiento humano»). Para empezar, dividía todas las percepciones mentales en dos categorías: «impresiones», que son o bien sensaciones externas o bien reflexiones internas, como los deseos, las pasiones y las emociones; e «ideas», que proceden de la memoria o de la imaginación. Argumentaba que las ideas, en última instancia, se copian de las impresiones, y utilizaba el término *conciencia* para referirse al pensamiento. Este principio de Hume, conocido como «principio de copia», es el primero en su ciencia de la naturaleza humana: «Todas nuestras ideas simples en su primera apariencia se derivan de impresiones simples que son correspondientes a ellas y que ellas representan exactamente». ⁵

Sin embargo —observa Hume—, nuestras ideas no surgen al azar; si así fuera no podríamos pensar coherentemente. En consecuencia, proponía un segundo principio, el «principio de asociación», según el cual es «imposible que las mismas ideas se unan regularmente en ideas complejas [...] sin que exista algún lazo de unión entre ellas, alguna cualidad que las asocie y por la que naturalmente una idea despierte a la otra». ⁶ Asimismo, estas asociaciones siguen tres principios: semejanza, contigüidad en tiempo y espacio, y causalidad. De estos, Hume reconoce que la causalidad nos lleva más allá de nuestros sentidos: vincula experiencias presentes a experiencias pasadas e infiere predicciones de futuro. El filósofo concluye que: «Todos nuestros razonamientos acerca de cuestiones de hecho parecen fundarse en la relación

de causa y efecto». ⁷ Esta conclusión predecía el declive del dualismo cartesiano.

Para Hume, la causalidad se compone de tres elementos fundamentales: prioridad en el tiempo, proximidad en el espacio y conexión necesaria. El filósofo sostiene que la idea de prioridad en el tiempo proviene de la observación y la experiencia. Así, si decimos que el acontecimiento A ha causado el acontecimiento B, eso significa que A ha precedido a B. La idea de proximidad en el espacio también proviene de la observación, puesto que, cuando observamos el acontecimiento B y decimos que ha sido causado por el acontecimiento A, es porque está próximo a él. Cuando salteo ajo en aceite de oliva, es mi casa la que de inmediato se llena de una deliciosa fragancia; no ocurre dos horas después, ni en casa de mis vecinos. El ajo salteado y la fragancia tienen que hallarse próximos entre sí, tanto en el tiempo como en el espacio, para que yo pueda inferir una relación de causa y efecto.

Para Hume, el problema lo planteaba el tercer componente: la conexión necesaria. El filósofo consideraba que la conexión necesaria entre causa y efecto no era algo que se derivara del mero pensamiento, o lo que él denominaba la «relación de ideas»; es decir, no era algo demostrablemente cierto, pero manifiesto independientemente de la experiencia, como $4 \times 3 = 12$. Lejos de ello, la conexión necesaria requería de la experiencia. Por ejemplo, puede que, mientras yo estoy salteando el ajo, suene el teléfono. ¿Concluiré que saltear ajo hace que suene el teléfono? No. Mi mente no inferirá ninguna conexión necesaria, a menos que eso ocurra cada vez que salteo ajo.

Imagínese que alguien le entregara unos polvos blancos que hasta ese momento desconocía. Usted no tendría ni idea de cuál podría ser el efecto de tragárselos. Aunque fuera un ingeniero aeroespacial, lo máximo que podría hacer sería describir el color, la textura, el olor y —en el caso de que fuera un ingeniero espacial especialmente imprudente, lo cual parece una contradicción en sus propios términos— el sabor. Pero sin la observación real o la

experiencia previa de los efectos de los polvos, no podrá conocerlos ni será capaz de predecirlos. Hume consideraba que la idea de una conexión necesaria entre causa y efecto se forma en la mente a partir de la experiencia; no es una característica real del mundo externo que podamos percibir a través de los sentidos. Y argumentaba que, cuando la gente considera que un conjunto de acontecimientos cualesquiera se hallan causalmente conectados, simplemente está observando que esos dos acontecimientos siempre van juntos: a los acontecimientos como A siempre les siguen acontecimientos como B; pero, como suele decirse, correlación no implica causalidad. El filósofo razona que, mediante la asociación, la impresión de un acontecimiento comporta la impresión del otro, y, si ambos continúan apareciendo juntos, a la larga la asociación se hace habitual. Entonces, cuando presenciamos el acontecimiento A, el hábito nos hace esperar el acontecimiento B. Esperamos la fragancia cuando salteamos ajo porque lo uno siempre sigue a lo otro, pero, en cambio, no esperamos que suene el teléfono. Aquí Hume prefigura a Pávlov y sus experimentos sobre condicionamiento. El filósofo concluye que, si este vínculo habitual entre los dos acontecimientos no es algo que pueda observarse a través de los sentidos, entonces la única fuente que podemos identificar como fundamento de la idea de causalidad es esa vinculación compulsiva que se produce en nuestro cerebro, la cual genera una expectativa; esa expectativa es la fuente de la idea de causalidad.

De modo que, basándonos en la experiencia repetida, inferimos o suponemos por la fuerza del hábito que, en el futuro, B siempre seguirá a A. Aquí nos tropezamos con una trampa lógica. Considere este ejemplo: es posible que usted haya comido gambas durante años, y luego, una noche, mientras conduce solo con su hijo de dos años, se detiene en un bar de carretera para tomar un refrigerio. Con avidez, hunde sus dientes en una succulenta gamba, y en cuestión de segundos nota cómo se le cierra la garganta, le cuesta respirar y se da cuenta de que está sufriendo una reacción alérgica. En lugar de la habitual placidez posprandial, le toca vivir un infierno. ¿Cómo

pasamos de «Comer gambas no me causa ningún problema» a «Esta noche comer gambas no me causará ningún problema», algo que normalmente funciona pero que en casos excepcionales (lo que el ensayista Nassim Taleb denominaría un «suceso del Cisne Negro») no lo hace? El problema es que, para hacer cualquier inferencia causal sobre el futuro, tenemos que partir de un supuesto: los efectos futuros serán como los pasados. Partimos de ese supuesto varias veces al día. ¿Cómo llegamos a él? Hume nos muestra que no lo hacemos a través de la lógica o la razón, porque en buena lógica es fácil concebir que los efectos futuros puedan ser distintos. Por lógica, usted sabe que puede intentar arrancar su coche, y, a diferencia de ayer y de todos los ayeres de los cinco años anteriores, descubrir que al girar la llave no ocurre nada. Hume también descarta nuestro uso de lo que él denomina «razonamiento probable» (empírico) debido a su carácter circular, ya que presupone lo mismo que se intenta probar (los efectos del futuro serán como los del pasado). Y concluye que nuestra creencia de que la naturaleza es uniforme y los efectos futuros serán como los pasados debe de generarse no de forma racional, sino psicológica, a partir de un hábito basado en la asociación.

Esto, sin duda, suponía adentrarse en un terreno peligroso, y Hume era consciente de ello. Estaba cuestionando nada menos que la idea de causalidad y dejando claro que esta era un axioma, una suposición carente de la esperanza de la prueba. No es que cuestionara que los efectos tuvieran causas, sino el origen de nuestra certeza de que así era. En 1754, en una carta a John Stewart, profesor de filosofía natural en Edimburgo, escribía: «Pero déjeme decirle que nunca he afirmado una proposición tan absurda, ya que cualquier cosa podría surgir sin una causa: yo únicamente sostenía que nuestra certeza de la falsedad de dicha proposición no procedía de la intuición ni de la demostración, sino de otra fuente».⁸

Hume mostró que, aunque transitar por el mundo resultaría extremadamente complicado sin él, el axioma de causa y efecto no puede probarse mediante las matemáticas, la lógica o la razón. A pesar de que admiraba a Newton, en este

caso puso en tela de juicio la base filosófica de la ciencia newtoniana y, con ella, la certeza mecanicista tanto en relación con el mundo como con la mente/cerebro. En los próximos capítulos veremos que ese cuestionamiento de la certeza mecanicista nos llevará a un territorio interesante.

Yo, yo mismo y... ¿quién?

Hume también tenía algunas cosas que decir sobre el yo, es decir, sobre nuestra identidad personal. Basándose en la introspección, concluía que el yo consiste en un conjunto de percepciones, pero que no hay un sujeto en el que dichas percepciones aparezcan: «Por mi parte, cuando penetro más íntimamente en lo que llamo mi propia persona, tropiezo siempre con alguna percepción particular de calor o frío, luz o sombra, amor u odio, pena o placer. No puedo jamás sorprenderme a mí mismo en algún momento sin percepción alguna, y jamás puedo observar más que percepciones». ⁹ Sin embargo, admite que debe explicar el propio hecho de que tiene una determinada idea de identidad personal, lo cual atribuye a asociaciones de percepciones. Para él, el yo no es más que un conjunto de experiencias vinculadas por las relaciones de causalidad y semejanza, formadas a partir de nuestra interminable cadena de percepciones. Nuestra imaginación las aglutina para generar una idea de identidad. Luego la memoria extiende esa idea más allá de las percepciones inmediatas, vinculándola con las del pasado.

En otras palabras, Hume pensaba que el origen de la idea de una identidad propia permanente se halla en nuestra forma de pensar, que confunde una sucesión de objetos relacionados entre sí, en este caso nuestras percepciones, con un objeto ininterrumpido e invariable dotado de identidad permanente, como una silla. Hume afirmaba: «Para justificarnos de este absurdo, fingimos frecuentemente algún nuevo principio ininteligible que enlaza estos objetos entre sí y evita su interrupción y variación. Así, fingimos la existencia continua de las percepciones de nuestros sentidos para evitar la interrupción y

recurrimos a la noción de un “alma”, “yo” y “substancia”, para desfigurar la variación. Sin embargo, podemos observar aún que, cuando no hacemos surgir esta ficción, nuestra propensión a confundir la identidad con la relación es tan grande que tendemos a imaginar algo desconocido y misterioso, que enlaza las partes, además de la relación». ¹⁰ Hume bien pudiera estar reprendiendo desde su tumba a los muchos que actualmente conciben la conciencia en esos términos.

Pero Hume también dirigía sus reproches a la *res cogitans*, o «cosa pensante», cartesiana. Él rechazaba la idea de que la mente fuera una cosa pensante; la veía más bien como un escenario donde el cerebro daba su espectáculo, «una especie de teatro donde varias percepciones aparecen sucesivamente, pasan, vuelven a pasar, se deslizan y se mezclan en una infinita variedad de posturas y situaciones». ¹¹ Hume no partía del supuesto de que una diversidad de experiencias se traduce en una sola unidad, en un sujeto. Razonaba, en cambio, que lo único que podemos captar de nosotros mismos mediante la introspección es un montón de percepciones e ideas, pero nunca captamos la mente que supuestamente las idea. Para él, el yo es un conjunto de percepciones sin un sujeto aglutinador, sin una esencia sustancial, persistente e inmutable. Si no hay un «yo» esencial, el dualismo de la sustancia es falso.

Dejamos el siglo XVIII con el que quizá fuera su mejor filósofo mirando con recelo a Descartes y Locke. Pero también a Hume se le escapaba algo importante: las numerosas mentes silenciosas de nuestro interior que influyen en nuestro comportamiento. Nuestra mente subconsciente, que nos controla a todos y cada uno de nosotros, es como un espía oculto en tu propia casa. Puede que nos parezca una fuerza sobrenatural, pero en realidad es el proceso neural que subyace al nivel de la percepción consciente.

Alemania y el nacimiento de la mente inconsciente

Cuando llegó el momento de que los alemanes participaran en el debate, estos

sacaron a colación el tema de los procesos mentales inconscientes. Por ejemplo, Arthur Schopenhauer escribía: «Porque el *Zeitgeist* de cada época es como un fuerte viento del este que lo invade todo. Puedes encontrar rastros de él en todo lo que se hace, se piensa y se escribe, en la música y la pintura, en el florecimiento de tal o cual arte: deja su huella en todo y en todos». ¹²

A comienzos del siglo XIX, a efectos de lo que aquí nos interesa, ese fuerte viento del este soplaba desde Alemania, y más concretamente desde la boca de Arthur Schopenhauer. Este privó del viento a las velas de Descartes, que creía que la mente era plenamente accesible y que nada se ocultaba a la reflexión consciente. Como filósofo, Schopenhauer se centró en las motivaciones del individuo, y llegó a la conclusión de que estas no eran demasiado elevadas: los humanos están motivados por su voluntad, y no por su intelecto, por más firmeza con que puedan negarlo. En su obra *El mundo como voluntad y representación*, publicada en 1818, concluía que «de hecho el hombre puede hacer lo que quiere, pero no puede desear lo que quiere». En esencia, no solo manda la voluntad (esto es, nuestras motivaciones subconscientes), sino que además el intelecto consciente ni siquiera se da cuenta de ello. Schopenhauer lo dejó claro al describir la voluntad como ciega y fuerte, y el intelecto como vidente pero paralítico: «el ejemplo más acertado para la relación entre ambos es el del ciego forzado que lleva en sus hombros a un paralítico que ve». ¹³

El enfoque de Schopenhauer situaba el problema de la conciencia en un terreno mucho más amplio. La mente, con todos sus procesos racionales, está muy bien, pero la clave es la «voluntad», que es la que nos da brío: «La voluntad [...] vuelve a llenar la conciencia de deseos, afectos, pasiones y cuidados». ¹⁴ Hoy en día los murmullos subconscientes de la «voluntad» siguen siendo insondables, y apenas se han hecho unas pocas incursiones en su territorio. Mientras escribo estas palabras, los entusiastas de la inteligencia artificial, cuyo objetivo es programar máquinas capaces de pensar como los humanos, han eludido e ignorado por completo este aspecto de la vida mental.

De ahí que David Gelernter, de la Universidad de Yale, uno de los mejores especialistas en ciencias informáticas del mundo, afirme que la agenda de la inteligencia artificial siempre se quedará corta: «Tal como existe actualmente, el campo de la inteligencia artificial no tiene nada que decir sobre las emociones y el cuerpo físico, por lo que simplemente se niega a hablar de ello». Gelernter sostiene que la mente humana incluye sentimientos además de datos y pensamientos, y que cada mente individual es un producto de las experiencias, las emociones y los recuerdos de esa persona en concreto, mezclados y remezclados a lo largo de toda una vida: «La mente se da en un cuerpo concreto, y la conciencia es obra de todo el cuerpo». Expresándolo en jerga informática, declara: «Puedo ejecutar una aplicación en cualquier dispositivo, pero ¿puedo ejecutar la mente de otra persona en tu cerebro? Obviamente no».¹⁵ Imagine, por ejemplo, al baloncestista Shaquille O'Neal y al actor y director Danny DeVito intercambiando sus cerebros: Danny se agacharía cada vez que atravesara una puerta, y Shaq fallaría los tiros por kilómetros.

La voluntad, para Schopenhauer, es voluntad de vivir, un impulso que engatusa a los humanos y a todos los animales para reproducirse. Para él, el propósito más importante de la vida humana es el producto final de la aventura amorosa, la descendencia, puesto que esta determina quién formará la próxima generación. En cambio, el filósofo sitúa al intelecto en el asiento trasero: no es el conductor del comportamiento, y tampoco está al tanto de las decisiones de la voluntad; es solo un portavoz tardío, inventando historias en su intento de explicar *a posteriori* lo que la voluntad ha forjado.

Al deponer al intelecto consciente, Schopenhauer también abrió la caja de Pandora del inconsciente. Describió nuestras ideas netamente conscientes como la mera superficie de un charco de agua, mientras que su interior estaba formado por sentimientos, percepciones, intuiciones y experiencias indistintas mezcladas con nuestra voluntad personal: «La conciencia es la mera superficie de nuestro espíritu, de la que, al igual que del globo terrestre, no conocemos el

interior sino solo la corteza». ¹⁶ Y afirmó que nuestro pensamiento real rara vez se daba en la superficie, y, por lo tanto, raramente podía describirse como una secuencia de «juicios claramente concebidos».

Schopenhauer introdujo el mundo de los procesos mentales inconscientes varias décadas antes de que Freud asumiera el protagonismo en ese ámbito, pero esta no era en absoluto una idea nueva. Recuérdese que Galeno había reconocido ya que muchos de los procesos corporales tienen lugar sin la intervención de la cognición, en particular los que mantienen vivo el cuerpo, como la respiración, y también nuestros impulsos naturales. Pero fue en el siglo XIX cuando la idea cobró impulso. En 1867, tras estudiar durante muchos años la fisiología del ojo, el médico, físico y filósofo de la ciencia materialista alemán Hermann von Helmholtz postuló que la inferencia inconsciente —es decir, un mecanismo involuntario, prerracional y de tipo reflejo— desempeña un importante papel en la percepción visual: el sistema visual del cerebro toma los datos visuales entrantes en bruto y los une en una imagen más coherente. ¹⁷ Se trata de un tipo de proceso distinto del que proponía Hume en su principio de copia, pero tampoco era una idea nueva, puesto que en el siglo XI el científico árabe Alhacén ya había sugerido algo similar.

Helmholtz fue mentor de su colega el médico y físico materialista Ernst Brücke. Ambos sostenían la idea de que los elementos que configuran la mente son físicos, y todas las relaciones causales entre dichos elementos se rigen por los mismos principios mecánicos que gobiernan la física y la química. Nada de espíritus vitales, ni de misticismo, ni de fantasmas. La mente y el cuerpo son uno. Más tarde Brücke se convertiría en profesor de fisiología en la Universidad de Viena, donde ejercería una gran influencia sobre uno de sus alumnos: Sigmund Freud. ¿Se imagina la intensa emoción de aquella atmósfera intelectual y científica? Se acabaron las presencias espectrales en el sistema: solo estaba el cerebro, compuesto por diversas partes, muchas de las cuales

funcionaban al margen de la percepción consciente, y todas ellas movidas por la química y la física.

En 1868, un oftalmólogo holandés llamado Franciscus Donders tuvo una idea que proporcionaría una nueva herramienta a todos los estudiosos del funcionamiento de la mente. Donders observó que, midiendo las diferencias en los tiempos de reacción visual de las personas, se pueden inferir diferencias en el procesamiento cognitivo, y sugirió que la cantidad de tiempo que se tarda en identificar un color determinado es la diferencia entre el tiempo que se tarda en reaccionar a dicho color en concreto y el tiempo necesario para reaccionar a la luz. Gracias a esta idea, los psicólogos comprendieron que podían estudiar la mente midiendo el comportamiento, y así nació el campo de la psicología experimental. De hecho, ese mismo método de Donders, junto con sus innovadoras ideas sobre el consumo de oxígeno cerebral, darían lugar a espectaculares avances en la comprensión de los procesos cognitivos mediante el uso de imágenes cerebrales, como harían por vez primera Marcus Raichle, Michael Posner y sus colegas de la Universidad de Washington en San Luis más de cien años después.

Todo aquel ruido de fondo en torno a la profunda mente inconsciente también se hizo sentir en Inglaterra, donde en 1867 era ya un concepto aceptado, como evidencian los escritos del psiquiatra británico Henry Maudsley: «La acción preconsciente de la mente, como la han denominado ciertos psicólogos metafísicos en Alemania, y la acción inconsciente de la mente, que actualmente ha quedado establecida más allá de toda duda racional, son ciertamente hechos que hasta el psicólogo introspectivo más ferviente debe admitir que la conciencia de uno mismo no puede explicarnos». ¹⁸ Maudsley pasaba luego a afirmar que «la parte más importante de la acción mental, el proceso esencial del que depende el pensamiento, es la actividad mental inconsciente». ¹⁹

Poco después, en 1878, inició su andadura la revista británica *Brain*. Al año siguiente la revista publicó un artículo del erudito Francis Galton en el

que este escribía sobre las conclusiones de un experimento que había realizado consigo mismo. Observaba una palabra escrita en una tarjeta y luego medía con un cronómetro cuánto tiempo tardaba en asociar dos ideas a dicha palabra, que luego anotaba. Tenía 75 palabras, y realizó el experimento en cuatro lugares muy distintos a intervalos de aproximadamente un mes. Los resultados le sorprendieron. A partir de su lista de 75 palabras, examinada cuatro veces, solo había producido 289 ideas diferentes. Casi el 25 % de las veces una palabra había producido los mismos términos asociados en las cuatro sesiones, mientras que en otro 21 % de las pruebas habían aparecido las mismas asociaciones en tres de los cuatro experimentos, lo cual mostraba mucha menos variedad de la que él esperaba. «Los caminos de nuestras mentes forman surcos muy profundos», señalaba Galton, para concluir:

Quizá la impresión más fuerte que me dejaron estos experimentos es la relativa a la diversidad del trabajo realizado por la mente en un estado de semiinconsciencia, y la razón válida por la que permiten creer en la existencia de estratos de operaciones mentales aún más profundos, sumergidos completamente por debajo del nivel de la conciencia, lo que puede dar cuenta de fenómenos mentales que de otro modo no podrían explicarse.²⁰

Al mismo tiempo, la mente consciente estaba a punto de obtener su propio campo de estudio. En 1874, un joven profesor de fisiología alemán llamado Wilhelm Wundt publicó el que sería el primer manual de psicología experimental, *Grundzüge der physiologischen Psychologie* («Principios de psicología fisiológica»), donde marcaba el territorio de la nueva disciplina, que incluía el estudio de los pensamientos, las percepciones y los sentimientos. A Wundt le interesaba especialmente analizar la conciencia, y pensaba que esta debía constituir el principal foco de atención de la psicología. Esbozó un sistema para investigar las experiencias inmediatas de la conciencia a través del autoexamen, que debía incluir una observación objetiva de los propios sentimientos, emociones, deseos e ideas. Cinco años después, en la Universidad de Leipzig, creó el primer laboratorio de psicología, ganándose así el calificativo de «padre de la psicología».

experimental». Wundt sostenía que la experimentación permitía identificar en la experiencia interior de los humanos regularidades similares a las de la ley; asimismo, creía que la neurofisiología y la psicología estudiaban el mismo proceso desde diferentes perspectivas: una desde dentro, y la otra desde fuera.

Freud, la mente inconsciente y sus coqueteos con el mecanicismo

Mientras tanto, la idea un tanto revolucionaria de la mente inconsciente empezaba a cobrar auténtico impulso con las aportaciones de Sigmund Freud. ¿Fue el factor de choque de las teorías psicoanalíticas de Freud el que las llevó a ocupar un primer plano? Fuera como fuese, lo cierto es que, en los comienzos de su carrera, Freud ya se mostraba audazmente ambicioso. En 1895 publicó *Proyecto de una psicología científica*, donde defendía la idea ultramaterialista de que a todo acontecimiento mental le correspondía un acontecimiento neurológico. Declaraba que el primer paso para alcanzar el objetivo de una psicología científica era identificar y describir con precisión el acontecimiento neuronal asociado a cada acontecimiento mental, lo que constituía una temprana versión de la actual búsqueda de los correlatos neuronales de la conciencia. Por si eso no bastara, a continuación proponía el segundo paso, el «reduccionismo eliminatorio»: había que eliminar el vocabulario utilizado para describir los estados mentales y reemplazarlo por un nuevo vocabulario neurológico. Así, por ejemplo, en lugar de hablar de sus celos, uno diría que su área J2 está funcionando a un determinado ritmo y velocidad concretos. Freud proponía ese cambio no solo para quienes estudiaran el cerebro, sino para todo el mundo. Absolutamente para todos. La poesía sería bastante distinta, al igual que las tarjetas de felicitación de San Valentín: «Mi p392J funciona un 95 % más rápido cuando mi L987T se corresponde con tu rostro». Probablemente el número de coordenadas relacionadas con ese «mi» habría ocupado demasiado espacio.

Pero, justo cuando su libro acababa de salir de imprenta, Freud cambió de opinión de manera absoluta y radical. Explica Owen Flanagan: «En 1895, el mismo año del *Proyecto*, Freud afirmaba que era una “mascarada inútil tratar de explicar fisiológicamente los procesos psíquicos”». ²¹ No solo eso; también decidió que solo se podía hablar de los acontecimientos mentales en el vocabulario de la psicología. Nada de reduccionismo. Flanagan sitúa una de las raíces de este argumento en el filósofo, psicólogo y exsacerdote alemán Franz Brentano, uno de los profesores de la Facultad de Medicina de Freud.

Brentano quería que la filosofía y la psicología se practicaran con métodos tan exigentes como los utilizados en las ciencias naturales. Diferenciaba entre dos enfoques distintos de la psicología: lo que él denominaba «enfoque genético», y otro enfoque, de carácter descriptivo. La psicología genética estudiaría la psique desde la perspectiva empírica tradicional, esto es, desde fuera, mientras que la psicología descriptiva, a la que a veces se refería como «fenomenología», pretendía describir la conciencia desde un punto de vista subjetivo, en primera persona. Coincidió con los filósofos del siglo XVIII en que todo el conocimiento se basaba en la experiencia, y argumentaba que la psicología debía emplear la introspección para estudiar empíricamente lo que uno experimentaba en su percepción interna. Aquí arraiga otra definición de la palabra *conciencia*: la percepción y la sensación subjetiva de la experiencia fenoménica, a la que volveremos en el próximo capítulo.

Brentano sostenía que la diferencia entre los fenómenos mentales y los fenómenos físicos es que estos últimos «son los objetos» de la percepción externa, mientras que los primeros se caracterizan por tener un contenido, por «hacer referencia» a algo, es decir, que «se dirigen a» un objeto. Ese objeto —especifica Brentano— «no debe interpretarse aquí como equivalente a una cosa», ²² sino que se trata de un objeto semántico. De modo que, igual que uno podría desear ver un caballo, también podría desear ver un unicornio, que es un objeto completamente imaginario; o podría desear el perdón, que, si bien puede o no ser imaginario, en cualquier caso es un objeto semántico que no

puede ponerse encima de la mesa. Brentano sostenía que ese «hacer referencia» era la principal característica de la conciencia, y aludía al estatus de los objetos del pensamiento con la expresión «inexistencia intencional». Escribe Owen Flanagan: «Esta perspectiva, que ha pasado a conocerse como “tesis de Brentano”, implica que ningún lenguaje que carezca de los recursos conceptuales necesarios para captar el contenido significativo de los estados mentales, como el lenguaje de la física o de la neurociencia, puede dar cuenta adecuadamente de los hechos sobresalientes relativos a los fenómenos psicológicos».²³

Basándose en sus experiencias con pacientes, y vinculándolas a la noción de procesos inconscientes, Freud elaboró una teoría psicológica sistemática. Dividió la mente en tres niveles: la «mente consciente», que incluye todo aquello de lo que tenemos consciencia; la «mente preconsciente», que contiene los recuerdos ordinarios, los cuales pueden recuperarse y remitirse a la mente consciente; y la «mente inconsciente», que es donde residen los sentimientos, impulsos, recuerdos y pensamientos que están fuera de la percepción consciente. La idea de que los procesos relacionados con las emociones, los deseos y las motivaciones son inaccesibles para la reflexión consciente no era nueva; de hecho, ya la habían planteado no solo Descartes, sino también, antes que él, Agustín de Hipona en el siglo IV y Tomás de Aquino en el XIII, y, después, Spinoza y Leibniz. Sin embargo, el planteamiento de Freud difería en que este consideraba indecorosos la mayoría de los contenidos del inconsciente; y según su teoría, el inconsciente influía en casi todos nuestros pensamientos, sentimientos, motivaciones, comportamientos y experiencias.

Curiosamente, Freud, que había defendido la idea de una psicología científica, nunca permitió que sus teorías psicoanalíticas se pusieran a prueba empíricamente en el naciente campo de la psicología experimental. Aunque algunas de las proposiciones de Freud han resistido el análisis empírico —por ejemplo, hoy se acepta de forma generalizada que la mayoría de los procesos cognitivos se realizan inconscientemente—, sus originales teorías en materia

de psicopatología no han aguantado un examen exhaustivo, y en general han ido a parar a la papelera.²⁴

Darwin lo cuestiona todo

Junto con la aceptación gradual de los procesos mentales inconscientes, en el siglo XIX surgió otra gran idea, esta vez fuera de las islas Británicas, con la publicación en 1859 de *El origen de las especies* de Charles Darwin. Las primeras ediciones volaron de las librerías y pronto despertaron el interés internacional. En la conclusión del libro, Darwin también se distanciaba del dualismo mente/cuerpo al escribir: «En el porvenir veo ancho campo para investigaciones mucho más interesantes. La psicología se basará seguramente sobre los [nuevos] cimientos [...] de la necesaria adquisición gradual de cada una de las facultades y aptitudes mentales. Se proyectará [...] luz sobre el origen del hombre y sobre su historia».²⁵ Aunque en un primer momento el libro de Darwin suscitó cierta desaprobación, para cuando se publicó *El origen del hombre*, en 1871, la gran teoría darwiniana de la evolución a través de la selección natural ya había sido ampliamente aceptada entre la comunidad científica y gran parte de la opinión pública en general. En este segundo libro, tras detallar numerosos ejemplos de la continuidad de los atributos físicos y mentales que comparten los animales y los hombres, concluía Darwin: «Por considerable que sea la diferencia entre el espíritu del hombre y el de los animales más elevados, es solo, ciertamente, una diferencia de grado y no de especie».²⁶ No es que estuviera dotando a los animales de un alma inmortal, sino que argumentaba de nuevo en contra del dualismo mente/cuerpo.

Como se ha señalado frecuentemente, Darwin era un hombre de modales exquisitos, y casi se disculpó por frustrar con su «mono» las creencias de tanta gente, incluida su propia esposa. De modo que concluyó *El origen de las especies* con una nota optimista y bastante esperanzadora para esas personas:

Así, la cosa más elevada que somos capaces de concebir, o sea la producción de los

animales superiores, resulta directamente de la guerra de la naturaleza, del hambre y de la muerte. Hay grandeza en esta concepción de que la vida, con sus diferentes fuerzas, ha sido alentada por el Creador en un corto número de formas o en una sola, y que, mientras este planeta ha ido girando según la constante ley de la gravitación, se han desarrollado y se están desarrollando, a partir de un principio tan sencillo, infinidad de formas, las más bellas y portentosas.²⁷

Darwin pensaba que su teoría también debía explicar las capacidades mentales humanas, pero esa parte de la teoría fue objeto de mayor controversia, y encontró resistencia tanto por parte de los dualistas tradicionales como de los empiristas seguidores de Locke y de Hume, los partidarios de la idea del cerebro humano como *tabula rasa*, que creían que todo el conocimiento proviene de la experiencia sensorial. Estas disputas asfixiarían los progresos para resolver el enigma de la conciencia durante muchos años. Finalmente acabaría tomando las riendas un enfoque distinto de la psicología; un enfoque que hundía sus raíces en los principios de asociación de Hume: el conductismo.

A finales del siglo XIX muchos filósofos insistían en la idea de que la mente debía de tener un cerebro físico, que de alguna manera contenía los recuerdos y la cognición. Algunos fisiólogos requerían asimismo la existencia de los nervios espinales, y otros sostenían que también el cuerpo en su conjunto formaba parte del paquete.

Locke separó la mente del alma y dotó a la primera de reflexión racional, acción ética y libre albedrío. La mente es la base de la conciencia, la volición y la personalidad, pero es falible y puede generar ilusiones y errores, y su única moneda son las ideas conscientes. Nada emerge desde las profundidades inconscientes. Locke esquivó el problema de cómo la materia podía producir algo como el libre albedrío añadiendo un Dios omnipotente a la ecuación y diciendo que él lo había hecho así. Hume eliminó esos poderes sobrenaturales de la ecuación e intentó establecer una verdadera ciencia de la mente humana. Al hacerlo, tomó conciencia de los límites de la mente, de cómo todo

pensamiento debía verse constreñido por las capacidades de esta, y, en consecuencia, cuestionó incluso la base filosófica de la ciencia mecanicista de Newton como forma de concebir el mundo, socavando los cimientos de la comprensión humana de la causalidad física.

Schopenhauer insistió en que lo que nos impulsa son las motivaciones e intenciones inconscientes, no el pensamiento consciente, lo que le convierte más bien en un apologista *a posteriori*. Helmholtz mostró que en realidad nuestros sistemas perceptivos no son meras fotocopiadoras, sino que aglutinan toda la información perceptual de la mejor manera posible. Luego vino Darwin, que metió nuestro cerebro en un continuo evolutivo, explicándonos cómo usar la selección natural para averiguar de qué modo ha llegado a ser como es y dejar a Dios al margen.

Entramos, pues, en el siglo XX todavía confundidos, todavía planteándonos las mismas preguntas, pero con un par de nuevas estrategias que podían emplearse: estudiar la diferencia en los tiempos de reacción a determinadas tareas, y utilizar una nueva psicología descriptiva centrada en una perspectiva subjetiva, en primera persona. Los siguientes cien años serían ricos en nuevas ideas, nuevos hallazgos científicos, formas totalmente nuevas de concebir la conciencia. La humanidad descifraría el átomo, descifraría el código del ADN, viajaría a la Luna y sería capaz de tomar imágenes del cerebro humano en vivo. Sin duda, habría algo que resolvería también el problema de la conciencia.

Los grandes progresos del siglo XX y los inicios del pensamiento moderno

Sin embargo, hay algunas personas —y yo soy una de ellas— que piensan que lo más práctico e importante en un hombre sigue siendo su visión del universo [...]. Creemos que la cuestión no es si la teoría del cosmos afecta a las cosas, sino si a la larga las afecta alguna otra cosa.

G. K. CHESTERTON

A comienzos del siglo XX, la filosofía de la mente y el cerebro todavía se hallaba dividida en dos bandos enfrentados: los racionalistas y los empiristas. Al finalizar el siglo, como veremos, las cosas no iban mucho mejor. Es casi como si nuestros cerebros humanos tuvieran un conjunto limitado de ideas, y, cualesquiera que sean los datos científicos o la atmósfera intelectual de la época, siempre acaba imponiéndose uno de esos dos puntos de vista. Pero volvamos a principios de siglo. Era el momento de que los recién llegados estadounidenses metieran baza sin miedo, y William James fue el primero en dar un buen repaso al tema de la conciencia. En 1907 ofreció una serie de conferencias en Harvard en las que empezó citando las palabras de G. K. Chesterton reproducidas más arriba, las cuales resumen perfectamente la gran pregunta de la filosofía en relación con la cuestión mente/cerebro: ¿puede un estado mental —una creencia inmaterial, una idea— afectar a la materia, esto es, a un estado cerebral?

James coincidía con Chesterton en que esa era la cuestión realmente importante. El tema de sus conferencias era un nuevo método filosófico, el pragmatismo, ideado por un amigo suyo, Charles Peirce, a partir de los debates que los dos hombres habían mantenido con otros filósofos y abogados en el Club Metafísico, una efímera pero influyente tertulia con sede en Cambridge, Massachusetts, de la que ambos habían sido cofundadores en la década de 1870. El pragmatismo no fue objeto de demasiada atención hasta que James lo desarrolló y promocionó veinte años después. En la primera de sus conferencias, James señaló algo que estaba oculto a simple vista; esto es: que los filósofos, junto con sus posturas filosóficas, estaban sesgados por su temperamento:

En gran medida, la historia de la filosofía es un choque entre temperamentos humanos [...] cualquiera que sea el temperamento del filósofo profesional, siempre que filosofa intenta enterrar su temperamento. Como el temperamento no es la razón convencionalmente reconocida, el filósofo solo esgrime razones impersonales para sus argumentos. Sin embargo, su temperamento le da una disposición más fuerte que cualquiera de sus premisas más objetivas. Hace que la evidencia pese para él en un sentido u otro, y favorece una visión del universo más sentimental o más dura de corazón, igual que lo haría cualquier hecho o un principio. El filósofo confía en su temperamento. Como desea un universo que se acomode a ese temperamento, cree en aquella representación del universo que se le acomode.¹

Y aquí está la parte auténticamente genial y descaradamente americana. James divide a los filósofos estadounidenses en dos grupos en función de su temperamento: «bostonianos de paso distinguido» y «tipos duros de las Montañas Rocosas»; una dicotomía temperamental que no solo detecta en la filosofía, sino también en la literatura, el arte, el Gobierno y los modales. Y, por supuesto, cada uno de estos dos grupos tiene en muy poca estima al otro: «Sus reacciones mutuas son como las que tienen lugar cuando los turistas bostonianos se mezclan con una población como la de Cripple Creek.² Cada tipo cree que el otro es inferior a él, pero en un caso el desdén está mezclado con la distracción, mientras que en el otro contiene una pizca de recelo». Luego perfila un poco más los dos grupos: los bostonianos de paso distinguido

y espíritu selecto son racionalistas (devotos de principios abstractos y eternos), intelectualistas, idealistas (en el sentido de que creen que todas las cosas proceden de la mente), optimistas, religiosos, defensores del libre albedrío, monistas (es decir, que el racionalismo parte de totalidades y universales, y da mucha importancia a la unidad de las cosas) y dogmáticos. Así que Descartes, en el fondo, es un «bostoniano».

Los espíritus rudos de las Montañas Rocosas son todo lo contrario: empiristas (se atienen a los hechos en toda su cruda variedad), sensacionalistas, materialistas (todas las cosas son materiales; no hay ninguna mente inmaterial), pesimistas, irreligiosos, fatalistas, pluralistas (los empiristas parten de las partes —valga la redundancia— y hacen del todo simplemente una colección de dichas partes) y escépticos (esto es, abiertos al debate). Hume resulta ser, pues, un espíritu rudo.

No obstante, James era consciente de que la mayoría de nosotros no somos estrictamente ni lo uno ni lo otro:

La mayoría de nosotros anhela las cosas buenas de uno y otro lado. Claro que los hechos son buenos: que se nos dé un montón de ellos. Pero los principios también son buenos: que se nos colme de principios. Indudablemente que el mundo es uno, si lo miras de cierta forma; pero indudablemente que es múltiple, si lo miras de otra. Es uno y múltiple, así que podríamos adoptar una especie de monismo pluralista. Todo se halla necesariamente determinado y, sin embargo, claro que nuestra voluntad es libre; así que la verdadera filosofía es un determinismo con libre albedrío. Es innegable la maldad de las partes, pero el todo no puede ser malvado: así que el pesimismo práctico puede combinarse con el optimismo metafísico. Y así sucesivamente. La gente corriente, lega en filosofía, nunca es radical, nunca sigue estrictamente su sistema, sino que habita vagamente en un sector u otro de ese sistema que sea plausible, uno que se acomode a las exigencias que le van surgiendo.³

En cambio, quienes tienen una mente más filosófica se sienten molestos por «tanta inconsistencia y vacilación en nuestro credo. No podemos conservar nuestra buena conciencia intelectual mientras sigamos mezclando cosas incompatibles de cada lado».

James define a los aficionados a la filosofía como personas que deseaban

hechos y ciencia, pero también religión, mientras que lo que encontraban era «una filosofía empírica que no es suficientemente religiosa, y una filosofía religiosa que no resulta suficientemente empírica». ⁴ Se necesitaba una ayuda práctica, y no una filosofía absolutista abstracta en extremo, para transitar por un mundo cuyos habitantes estaban interesados en la ciencia con la que se los estaba bombardeando, pero también encontraban consuelo en la religión o el romanticismo. James creía que el método pragmático proporcionaría esa ayuda. Su fundamento era la idea de que nuestras creencias son también nuestras reglas con vistas a la acción; que, cuando adoptamos una creencia, adquirimos la predisposición a actuar de una cierta manera. Para comprender la importancia de una creencia, basta determinar qué acción produciría dicha creencia. Si dos creencias distintas producen la misma acción, más vale dejarlo correr:

En primer lugar, el método pragmático es un método para resolver disputas metafísicas que de otra manera podrían resultar interminables. ¿Es el mundo uno o múltiple? ¿Libre o determinado? ¿Material o espiritual? En esta serie de preguntas, cada una de las dos alternativas puede considerarse apropiada o no sobre el mundo. Desde luego, las disputas que suscitan no tienen fin, y en tales casos el método pragmático trata de interpretar cada una de esas ideas señalando sus respectivas consecuencias prácticas. ¿Qué diferencia de orden práctico supondría para alguien el que fuera verdadera tal idea en vez de su contraria? Si no puede señalarse ninguna diferencia práctica, entonces las alternativas significan lo mismo de manera práctica, y toda disputa es vana. Cuando una disputa es seria, debemos ser capaces de mostrar alguna diferencia práctica que tiene que seguirse si un lado o el otro está en lo cierto. ⁵

Aunque el pragmatismo se basa en la idea de que un estado mental podría ser la causa de una acción, es solo un método y no aboga por determinados resultados concretos. Está abierto a diversos tipos de metodologías, que se emplean en diferentes ciencias. Sin embargo, es un método que *a priori* rechaza la metafísica y el intelectualismo basado en cantidades interminables de pensamiento. Como tal, resultó especialmente atractivo para los profesionales de la psicología de estímulo-respuesta —seguidores de las teorías de asociación de Hume— que dominaban el nuevo campo de la

psicología experimental inaugurado por Wilhelm Wundt; asimismo, un alumno de Wundt, el carismático psicólogo Edward Titchener, desarrolló el método y lo llevó a Nueva York. Otro personaje especialmente influyente fue Edward Thorndike. En su monografía de 1898 *Animal Intelligence: An Experimental Study of the Associative Processes in Animals* («Inteligencia animal: estudio experimental de los procesos asociativos en los animales»), formuló el que sería el primer enunciado general sobre la naturaleza de las asociaciones: la denominada «ley del efecto». Thorndike había observado que una respuesta a la que seguía una recompensa dejaba una impronta en un organismo que la convertía en una respuesta habitual, y que dicha respuesta desaparecía si no se daba recompensa alguna. Este mecanismo de estímulo-respuesta podía ser el responsable de crear respuestas cada vez más adaptativas.

Pronto la psicología de estímulo-respuesta, también conocida como conductismo, llegó a dominar el estudio de los procesos asociativos en Estados Unidos. Los conductistas abordaban la psicología partiendo de la perspectiva de que su objeto de estudio apropiado era el comportamiento antes que la experiencia mental y subjetiva; y este debía estudiarse utilizando los métodos propios de las ciencias naturales, no la introspección. Consideraban que, dado un determinado estímulo ambiental, el comportamiento de cualquier animal, incluidos los humanos, podía explicarse por una tendencia a reaccionar de cierta manera que a su vez funcionaba de manera parecida a una ley.

Dominaba este campo del saber la dinámica figura de John B. Watson. La postura de Watson era que la psicología solo podía ser objetiva si se basaba en un comportamiento observable, y rechazaba toda la palabrería sobre procesos mentales que no pudiera observarse de manera pública y notoria; de hecho, estaba vetado mirar en el interior de la caja negra del cerebro. Ignorando la teoría darwiniana de los procesos mentales innatos, Watson se aferró a la idea de que todo el mundo está dotado exactamente del mismo equipamiento neuronal; la mente es una tabla rasa, y se puede entrenar a

cualquier niño para hacer cualquier cosa aprendiendo mediante estímulo-respuesta y recompensa. Esta idea resultaba especialmente atractiva para la concepción estadounidense de la igualdad entre todas las personas, de modo que la mayoría de los directores de departamentos de psicología de Estados Unidos no tardaron en sustentar estos puntos de vista, ignorando el postulado de la teoría darwiniana de que la complejidad está integrada en el organismo humano a través del proceso de selección natural y evolución. El conductismo reinaría en Estados Unidos durante las cinco décadas siguientes, presidido durante muchos años por el que sería su principal portavoz, el profesor de psicología de Harvard B. F. Skinner.

Obviamente, incluso en las épocas en las que el mundo académico parece estar dominado por grandes temas, existen corrientes opuestas. De hecho, se estaban desarrollando nuevos métodos para estudiar los procesos «mentales» que no solo se iban incorporando constantemente a la psicología experimental, sino que además llegarían a convertirse en las herramientas de exploración dominantes en los tiempos modernos.⁶ Aun así, hablar de estados mentales y de conciencia seguiría estando fuera de lugar en Estados Unidos hasta el surgimiento, a mediados de siglo, de la revolución cognitiva liderada por George A. Miller en Harvard y la noción de mentalismo formulada por Roger W. Sperry en el Instituto de Tecnología de California (Caltech).

La resistencia de los canadienses y el auge de la neurociencia moderna

Por fortuna, los investigadores de Canadá no se subieron al carro conductista. De hecho, el primer neurocirujano de Montreal, Wilder Penfield, estaba realizando descubrimientos sorprendentes en pacientes con convulsiones que solo podían controlarse eliminando la parte de la corteza cerebral que las provocaba. A fin de localizar la región de los focos convulsivos, Penfield estimulaba diferentes partes de la corteza cerebral con sondas eléctricas y

observaba las respuestas del paciente. Durante la intervención, los pacientes estaban despiertos (se les administraba únicamente anestesia local) para que pudieran explicar lo que sentían o dejaban de sentir. De este modo, Penfield elaboró mapas de las cortezas sensorial y motora que correspondían a las diversas partes del cuerpo; es decir, una representación física del cuerpo humano ubicada en el cerebro.⁷ Sin embargo, normalmente el cuerpo no se representaba de manera proporcional, sino que la proporción respondía al grado de inervación de cada una de sus partes: cuanto mayor era la inervación de una parte determinada, mayor era el área cerebral dedicada a ella. Penfield y su estrecho colaborador, el fisiólogo Herbert Jasper, fueron los primeros en entender la localización de la función cerebral. Escribía Penfield: «La conciencia continúa, independientemente de qué área de la corteza cerebral se elimine. Por otra parte, la conciencia se pierde inevitablemente cuando la función del tronco cerebral superior (diencefalo)⁸ se ve interrumpida por una lesión, presión, enfermedad o descarga epiléptica local». Sin embargo, se apresuraba a matizar que «sugerir que existe un bloque del cerebro donde se localiza la conciencia sería como hacer volver a Descartes y ofrecerle un sustituto de la glándula pineal como sede del alma».⁹

Luego Penfield pasaba a explicar que, si bien la información sensorial se procesa a través del diencefalo, es decir, a través de las regiones subcorticales, dicha información va y viene entre la subcorteza y las diferentes áreas de la corteza: «Así, los procesos de diferenciación de la mente son posibles gracias a la actividad funcional combinada en el diencefalo y la corteza cerebral, y no solo en el diencefalo».¹⁰ También afirmaba que el proceso final necesario para la experiencia consciente es el de centrar la atención en el estado mental que la produce, y predecía que ese proceso forma parte de la función del diencefalo. Podemos discernir en estos textos que Penfield utiliza la palabra *conciencia* para aludir a dos cosas distintas: en el primer caso está hablando del estado mental de estar despierto y consciente, es decir, no en coma; en los otros dos, en cambio, se refiere a la conciencia en

el sentido de Descartes, es decir, a un pensamiento, o un pensamiento sobre un pensamiento, y es aquí donde añade que centrar la atención es un componente necesario.

Penfield añadió a su grupo a un psicólogo, Donald Hebb, para estudiar los efectos que producía en sus pacientes una lesión cerebral, así como los resultados de la cirugía en el funcionamiento del cerebro. Como la mayoría de los profesionales que han estudiado a pacientes con lesiones cerebrales, Hebb terminó convencido de que el funcionamiento del cerebro explicaba el comportamiento. Aunque actualmente esto puede parecernos algo elemental, como se lo pareció a Galeno hace tantos siglos, en 1949 —cuando Hebb publicó su libro *The Organization of Behavior: A Neuropsychological Theory* («La organización del comportamiento: una teoría neuropsicológica») — el dualismo mente/cuerpo todavía cautivaba a muchas personas. En una época en la que la psicología se hallaba aún bajo el firme dominio del conductismo, Hebb tomó por asalto el mundo psicológico al irrumpir audazmente en la caja negra del cerebro ignorando los «vetos» impuestos tanto por el empirista Hume como por los conductistas. Formuló la hipótesis de que múltiples neuronas podían combinarse en una coalición, dando lugar a una única unidad de procesamiento. Las pautas de conexión de tales unidades, que podían cambiar, constituían los algoritmos (también susceptibles de cambiar al hacerlo las pautas de conexión) que determinaban la respuesta del cerebro a un estímulo. De esta idea surgiría el mantra «Las células que se activan a la vez se conectan entre sí».¹¹ Según esta teoría, el aprendizaje tiene una base biológica en las pautas de interconexión de las neuronas. Hebb señalaba que el cerebro se mantiene constantemente activo, no solo cuando se lo estimula, y los estímulos externos solo pueden modificar esa actividad en curso. La propuesta de Hebb tenía sentido para quienes diseñaban redes neuronales artificiales, y empezó a utilizarse en programas informáticos. Al abrir la caja negra del cerebro y mirar en su interior, Hebb también lanzó la andanada inicial que desencadenaría la revolución contra el conductismo.

La revolución cognitiva en Estados Unidos

El dominio del conductismo sobre la psicología estadounidense empezó a declinar en la década de 1950, especialmente cuando un grupo de jóvenes y brillantes investigadores de la mente, como Allen Newell, Herbert Simon, Noam Chomsky y George Miller, se unieron para fundar la psicología cognitiva. Miller, por ejemplo, hizo lo que se supone que tiene que hacer un científico cuando se le presenta una evidencia convincente: cambiar de opinión. Estaba investigando el habla y la audición en Harvard cuando escribió su primer libro, *Lenguaje y comunicación*. William James habría quedado impresionado por el prefacio, donde Miller declaraba abiertamente su parcialidad: «El sesgo [de la obra] es conductista». En la sección sobre psicología, que trataba de las diferencias en el modo como las personas utilizan el lenguaje, su modelo probabilístico de selección de palabras se basaba en una pauta conductista de aprendizaje por asociación. El título de un libro de texto que escribiría once años después, *Psychology: The Science of Mental Life* («Psicología: la ciencia de la vida mental»), anunciaba un cambio radical con respecto a su postura anterior de que la psicología debía estudiar únicamente el comportamiento. Lo que provocó el cambio de opinión de Miller fue el auge de la teoría de la información: la introducción del denominado «Lenguaje de Procesamiento de la Información I», un lenguaje informático que implementaba varios programas de inteligencia artificial previos, y las ideas sobre la organización neuronal del genio informático John von Neumann, que postulaba que el cerebro podía funcionar de manera similar a un enorme ordenador de computación paralela; esta última implica la posibilidad de ejecutar varios programas al mismo tiempo, en contraposición a la programación secuencial, en la que solo se puede ejecutar un programa a la vez.

Es posible que, para Miller, el golpe de gracia al conductismo fuera su encuentro con el brillante lingüista Noam Chomsky, que por entonces estaba sacudiendo hasta las raíces el mundo psicológico al mostrar que la

predictibilidad secuencial del habla seguía reglas gramaticales, y no reglas probabilísticas; además, dichas reglas gramaticales resultaban ser de lo más chocantes, puesto que eran innatas y universales, es decir, que todo el mundo las tenía y estaban ya «grabadas» en el cerebro al nacer. De ese modo no hubo más remedio que descartar la noción de *tabula rasa*, aunque no sin una fuerte patalota cuyos gritos aún se oyen.

En septiembre de 1956, Chomsky publicó un ensayo titulado «Tres modelos para la descripción del lenguaje», su versión preliminar de aquellas ideas sobre las teorías sintácticas. Tomando el mundo lingüístico por asalto, con aquel trabajo transformó de un plumazo el estudio del lenguaje. La conclusión que extrajo Miller del ensayo fue que el asociacionismo, la mascota de los conductistas —y especialmente del conductista radical B. F. Skinner—, no podía explicar cómo se adquiere el lenguaje. Si bien los conductistas habían dilucidado algunos aspectos del comportamiento, había algo más en aquella caja negra que estos no habían podido ni podrían explicar nunca. Ya era hora de que alguien intentara descifrarlo.

Miller empezó a explorar las implicaciones psicológicas de las teorías de Chomsky con el propósito último de comprender cómo funcionan el cerebro y la mente como un todo integrado. Por entonces, no obstante, había un aspecto de la vida mental que le provocaba cierto recelo, y que le llevó a escribir en su *Psychology* que por el momento había que aparcar su estudio: «*Conciencia* es una palabra desgastada por un millón de lenguas. Dependiendo de la figura retórica elegida, es un estado del ser, una sustancia, un proceso, un lugar, un epifenómeno, un aspecto emergente de la materia o la única realidad verdadera. Tal vez deberíamos prohibir la palabra durante una o dos décadas hasta que podamos desarrollar términos más precisos para los diversos usos que actualmente oculta *conciencia*». ¹²

El término *conciencia*, que Descartes empleara para aludir tanto a un pensamiento como a un pensamiento sobre un pensamiento, había florecido con los años y adoptado toda clase de significados adicionales. Además de lo

que escribió Miller, su significado también había pasado a interrelacionarse con conceptos como la percepción, la autopercepción, el autoconocimiento, el acceso a la información o la experiencia subjetiva. Aunque la mayoría de los investigadores siguieron el consejo de Miller y aparcaron el estudio de la conciencia, hubo un grupo especialmente intrépido de ellos que no lo hicieron. Lejos de ello, se dedicaron a realizar un inventario de lo que la ciencia podía decir sobre la conciencia hasta aquel momento.

En busca de claridad en el Vaticano

En 1964, mientras Miller cerraba bajo llave la palabra *conciencia*, la Pontificia Academia Scientiarum (Pontificia Academia de las Ciencias) le asignaba un papel protagonista como centro de atención de una semana de estudio. Esta institución vaticana tiene sus raíces en la denominada Accademia dei Lincei (Academia de los Linceos), fundada en 1603 por un príncipe y naturalista romano de dieciocho años llamado Federico Cesi, cuyo tío era un cardenal muy bien relacionado. Cesi la fundó con el propósito de entender las ciencias naturales a través de la observación, la experimentación y el razonamiento inductivo; y para simbolizar dichos objetivos eligió como emblema al linceo, por la especial agudeza de su visión. En 1610, Galileo fue nombrado director de la institución.

Pero aquella era una época difícil para tales aventuras, y la academia no sobrevivió a la temprana muerte de Cesi a los cuarenta y cinco años de edad. No obstante, en 1847 fue refundada por el papa Pío IX con el nombre de Pontificia Accademia dei Nuovi Lincei (Academia Pontificia de los Nuevos Linceos). Posteriormente, tras la unificación de Italia y su separación del Vaticano en 1870, la Academia de los Nuevos Linceos se dividió en dos: la Real Academia Nacional Linceana, bajo bandera italiana, y una nueva institución que a la larga se convertiría en la Pontificia Academia de las Ciencias, refundada en 1936 por el papa Pío XI y con sede en la Ciudad del

Vaticano. Pese a su fundación papal y al hecho de estar ubicada dentro de los muros de los Jardines Vaticanos, esta última no plantea restricción alguna en cuanto a sus investigaciones. La integran científicos de numerosos países y disciplinas, y tiene como uno de sus objetivos «fomentar el progreso de las ciencias matemáticas, físicas y naturales, y el estudio de las cuestiones epistemológicas relacionadas». En septiembre de 1964, la Pontificia Academia organizó una semana de estudio sobre el tema «Cerebro y experiencia consciente», que dirigiría el renombrado médico y fisiólogo sir John Eccles.

Eccles era australiano. Durante su estancia en la Facultad de Medicina, no solo había sido un estudiante despierto, sino que también había practicado el salto con pértiga. Leer *El origen de las especies* para su clase de zoología le incitó a leer también textos filosóficos, tanto clásicos como contemporáneos, sobre la cuestión mente/cerebro.¹³ Pero la Facultad de Medicina no podía responder a sus preguntas sobre la interacción entre mente y cuerpo, de modo que decidió hacerse neurocientífico.¹⁴ También se propuso obtener una Beca Rhodes para ir a Oxford y trabajar con el afamado neurofisiólogo Charles Sherrington. Y lo hizo. En 1925 partió rumbo a Inglaterra, a medio mundo de distancia.

Eccles empezó a estudiar el método de transmisión neuronal en la sinapsis. Al principio estaba convencido de que dicha transmisión era de naturaleza eléctrica. En ese período conoció al filósofo Karl Popper, que le alentó a comprobar su hipótesis con el máximo rigor en consonancia con su idea de que la fuerza de una hipótesis dependía, no de las evidencias que aparentemente la sustentaban, sino del fracaso de cualquier investigación exhaustiva que pretendiera probar su falsedad. Tras someter pertinazmente su hipótesis a diversas pruebas, Eccles cambió de opinión y decidió que la transmisión sináptica era de naturaleza química. Este cambio radical llevó a escribir a un viejo amigo suyo, el fisiólogo sir Henry Dale: «¡Qué extraordinaria conversión! Uno recuerda, casi de manera inevitable, a Saulo en su camino a

Damasco, cuando de repente brilló aquella luz y se le cayó la venda de los ojos». ¹⁵ Durante la década siguiente, Eccles se dedicó a dilucidar los mecanismos implicados en la activación e inhibición de la sinapsis de la neurona motora en la médula espinal, y luego pasó a centrar su atención en el tálamo, el hipocampo y el cerebelo. El año anterior al congreso en la Pontificia, Eccles fue galardonado con el Premio Nobel de Fisiología y Medicina, y unos años antes había recibido el título de sir por la misma investigación. En su época fue toda una leyenda, y quienes le conocimos fuimos testigos de su chispeante inteligencia, su energía inagotable y sus grandes cualidades como científico. Por otra parte, se había educado en el catolicismo y era un dualista declarado. Al pragmatista William James no le habría sorprendido saber que su creencia condicionaba su acción y que pasó toda su vida buscando los mecanismos mediante los cuales la mente controla el cuerpo.

En un ensayo publicado en 1951 en la revista *Nature* que llevaba por título «Hipótesis relativas al problema cerebro-mente», Eccles afirmaba que «muchos hombres de ciencia encuentran en el dualismo y la interacción los postulados iniciales más aceptables con vistas a un enfoque científico del problema de la mente y el cerebro. En dicho enfoque surge la pregunta: ¿qué hipótesis científicas cabe formular que guarden alguna relación con el problema hasta ahora refractario de la relación cerebro-mente?». ¹⁶ Y a continuación pasaba a postular tal hipótesis. Aunque él pensaba que toda experiencia perceptiva es el resultado de una pauta específica de activación neuronal, y que la memoria obedece a un aumento de la eficacia sináptica, por alguna razón creía que la experiencia y la memoria son «inasimilables al sistema materia-energía». En su lugar, proponía que la corteza activa poseía «una sensibilidad de un tipo distinto de la de cualquier instrumento físico», y que «la mente se vincula al cerebro ejerciendo campos de influencia espaciotemporales que devienen efectivos a través de esta única [...] función de la corteza cerebral activa». ¡Vaya! Eso viene a ser prácticamente vudú con

un lenguaje sofisticado. Con ello, Eccles reemplazaba la glándula pineal de Descartes por una corteza cerebral activa misteriosamente sensible. De hecho, doscientos años después de Descartes, daba continuidad a su tradición dualista pese a dedicar sesenta horas a la semana a trabajar con neuronas y registrar sus reacciones, y pese a que, por lo demás, había adoptado por completo la agenda determinista. ¡Alucinante!

Parte de la labor de Eccles como director de la semana de estudio consistía en seleccionar a los demás asistentes y publicar los debates, lo que se traduciría en una obra de gran trascendencia titulada *Brain and Conscious Experience* («El cerebro y la experiencia consciente»). El único sesgo del que se podía acusar a Eccles era de haber sobrecargado ligeramente el congreso de fisiólogos, pero todos ellos tendían a ser polivalentes. El caso es que logró reunir a los mejores científicos de sus respectivos campos, que abarcaban la neurofisiología, la neuroanatomía, la psicología, la farmacología, la patología, la biopsicología, la neurocirugía, la química, las comunicaciones, la cibernética, la biofísica y el comportamiento animal. La Academia, fiel a su objetivo de estudiar las ciencias físicas, matemáticas y naturales, había declarado una única restricción: nada de filósofo. A Eccles eso no le hacía ninguna gracia, pero el grupo incluía lo que el autor de una reseña sobre la obra calificaría de «filósofos aficionados de no poco nivel», para concluir más adelante que, «como único volumen que trata de los recientes progresos en nuestra comprensión de la corteza cerebral, probablemente [*Brain and Conscious Experience*] no tiene parangón». ¹⁷

Antes de la reunión, los participantes recibieron un informe de la Academia en el que se describía la conciencia como «el concepto psicofisiológico de la capacidad perceptiva, de la consciencia de la percepción, y la capacidad de actuar y reaccionar en consecuencia». Como me comentó Roger Sperry, mi mentor y futuro premio Nobel, a su regreso al Instituto de Tecnología de California, «el papa dijo: “El cerebro es vuestro, pero la mente es nuestra”».

Los debates se dividieron más o menos entre estos tres aspectos de la conciencia: percepción, acción y volición.

El zoólogo del grupo, William Thorpe, desarrolló la idea:

El término conciencia, aunque tiene innumerables connotaciones de significado, implica, creo, tres componentes básicos. En primer lugar, una consciencia de sensibilidad interior, lo que podría denominarse «tener percepción interna». En segundo término, una consciencia del yo, de la propia existencia. En tercer lugar, la idea de conciencia incluye la de unidad; es decir, implica en un sentido bastante vago la fusión de la totalidad de las impresiones, pensamientos y sentimientos que configuran el ser consciente de una persona en un todo único.¹⁸

Al hablar de la relación de los acontecimientos cerebrales con la experiencia consciente, Eccles formuló esta pregunta: «¿Cómo puede una pauta espaciotemporal concreta de actividad neuronal en la corteza cerebral evocar una determinada experiencia sensorial?». ¹⁹ La pregunta no encontró respuesta, y sigue sin encontrarla.

Al leer el mencionado volumen, no puedo menos que esbozar una sonrisa ante el impacto que tuvo la charla de Roger Sperry en torno a nuestra investigación sobre la operación conocida como «cerebro escindido» o callosotomía,²⁰ que por entonces estaba dando sus primeros pasos. En su resumen escrito, Sperry había declarado: «Todo lo que hemos visto hasta ahora indica que la cirugía ha dejado a estas personas con dos mentes distintas, es decir, dos esferas de conciencia distintas». ²¹ La animada discusión que se produjo a continuación revela lo fascinantes que resultaban nuestros hallazgos y la sensación que causó la charla de Sperry. Este les estaba diciendo al Vaticano y a sus colegas que la mente podía dividirse en dos con el corte de un bisturí.

Por entonces Sperry estaba en pleno proceso de cambio de opinión —en parte debido a la propia investigación sobre el cerebro escindido— y de reajuste de su postura básica en torno a la función cerebral. Empezaba a dar la espalda al materialismo y al reduccionismo tal como se definían entonces y a calificarse a sí mismo de «mentalista». Anteriormente, aquel mismo año,

mientras trabajaba en una conferencia de divulgación sobre la evolución del cerebro, se sorprendió a sí mismo concluyendo que «las facultades mentales emergentes deben ejercer lógicamente un control causal descendente sobre los acontecimientos electrofisiológicos de la actividad cerebral». ²² En aquel momento, la noción de que el estado mental podía influir en un estado cerebral era una auténtica herejía en el mundo de la neurociencia, y en gran medida todavía sigue siéndolo. También yo llegué a conclusiones similares y reintroduje la idea de que los procesos mentales tienen un efecto causal descendente en una charla que di en Edimburgo en 2009, en el marco de las Conferencias Gifford, para descubrir, una vez más, que los deterministas de toda laya no se mostraban demasiado receptivos a la idea. La premisa central tanto del conductismo como del materialismo es que el proceso cerebral físico objetivo es una red de estímulo-respuesta causalmente completa en sí misma; es decir, que no recibe ningún estímulo externo procedente de fuerzas conscientes o mentales, ni tampoco lo necesita. En muchos sentidos, el presente libro constituye un nuevo intento de lidiar con este problema.

En el congreso del Vaticano, Sperry minimizó su creciente postura mentalista al limitarse a concluir que «la conciencia puede tener un valor operativo real, que es algo más que un mero matiz, un subproducto, un epifenómeno o un paralelismo metafísico del proceso objetivo». ²³ En otro lugar parafraseaba esto mismo hablando de «una visión que sostiene que la conciencia puede tener algún uso operativo y causal». ²⁴

Eccles, mostrando su lado materialista, admitió: «Estoy dispuesto a afirmar que, como neurofisiólogos, simplemente no necesitamos para nada la conciencia en nuestras tentativas de explicar cómo funciona el sistema nervioso», ²⁵ reconociendo a un tiempo: «Obviamente, no me creo esa historia; pero a la vez tampoco conozco la respuesta lógica a ella». ²⁶ Mantenía, pues, su posición dualista.

Al final de la semana, la tarea de recapitular las conclusiones del congreso recayó en el psicólogo del Instituto de Tecnología de Massachusetts (MIT)

Hans-Lukas Teuber, uno de los fundadores de la neuropsicología, famoso por sus brillantes «sumarios» de congresos y reuniones científicas, que aderezaba con elaborados movimientos de cejas.²⁷ En su intervención, Teuber hizo un gráfico representando los puntos en los que los participantes coincidían y discrepaban, y señaló dónde había lagunas de conocimiento, proporcionando un conciso resumen de la situación vigente de este campo de investigación como solo él podía hacerlo. Los demás coincidieron en que sabían bastante sobre el procesamiento cortical de la sensación y la visión, y que, si llegaban a obtener un conocimiento equiparable sobre la acción motora, la retención de memoria y la consciencia —que no era el caso—, habrían avanzado mucho con vistas a entender la experiencia consciente. No obstante, Teuber se lamentó: «Al tratar de delinear los sistemas o mecanismos que podrían ser necesarios para la consciencia, parece que hayan surgido todas las divergencias de opinión concebibles. Ni siquiera estábamos seguros [...] del modo de determinar para qué sirve la consciencia».²⁸

Teuber era un hombre apasionado que me ayudó haciendo las funciones de mentor en los comienzos de mi carrera. Recuerdo su visita a Santa Bárbara (California) a finales de la década de 1960. Mi esposa y yo habíamos organizado una recepción para él en nuestra casa, en la pequeña población californiana de Mission Canyon, cuando se me acercó, me guiñó un ojo y me dijo que quería hablar conmigo a solas. Nos dirigimos al dormitorio; entonces sacó de su maletín un manuscrito que yo había enviado recientemente a la revista *Neuropsychologia* y empezó a corregirlo con un lápiz rojo. Yo me quedé perplejo, aunque a la vez me sentí agradecido por la atención que me dispensaba. Cuando terminamos, se levantó de un salto y exclamó: «¡Volvamos a la fiesta!». Yo debí de decir algo coherente, porque a continuación me invitó a unirme al recién fundado Simposio Internacional de Neuropsicología, una maravillosa organización cuyos integrantes se reúnen anualmente en diferentes ciudades del mundo (un evento del que tuve la oportunidad de disfrutar durante veinte años).

Evidentemente, el congreso del Vaticano no resolvió el problema mente/cuerpo, pero las diversas opiniones allí expresadas desencadenaron toda una serie de argumentaciones y debates en el seno de la biología y la filosofía que se prolongarían hasta hoy. Sobre la mesa seguía estando la misma lista de soluciones posibles, con Eccles aferrado a la visión cartesiana de que la mente y el cuerpo eran dos entidades distintas por más que nunca pudiera encontrar evidencias empíricas de ello. La mayoría, no obstante, se inclinaba hacia la visión materialista de que la mente —la conciencia— era un producto de la materia, aunque el modo como eso sucedía seguía siendo tan enigmático como siempre.

El congreso del Vaticano supuso un punto de inflexión para Sperry. La posibilidad de que los estados mentales pudieran afectar causalmente a los estados cerebrales se convirtió en su pasión científica, con todas sus consecuencias. El psiquiatra del grupo, Hans Schaefer, de la Universidad de Heidelberg, suscribía también esa teoría basándose en su creencia de que el psicoanálisis realmente funcionaba. Por otra parte, las teorías evolutivas permitieron que las teorías materialistas de la conciencia revistieran dos formas distintas, que pasarían a conocerse como «emergentismo» y «pampsiquismo». La primera postula que la conciencia emerge de la materia inconsciente una vez que esta alcanza un cierto nivel de complejidad u organización. Sperry se inclinaba en ese sentido. La segunda, el pampsiquismo, elimina el problema de un plumazo al sugerir que toda materia tiene conciencia subjetiva, aunque de una amplia gama de tipos distintos. La idea en este caso es que no hacen falta para nada los conceptos de emergencia y complejidad para explicar la conciencia, ya que esta sería una característica primordial de todas las cosas, desde las rocas y las hormigas hasta nosotros.

A su regreso del congreso, Sperry siguió refinando su punto de vista, y al año siguiente salió del armario mentalista en una conferencia pronunciada en su *alma mater*, la Universidad de Chicago: «Voy a alinearme en una postura contraria, junto con esa minoría mentalista de aproximadamente el 0,1 %, en

apoyo de un modelo cerebral hipotético en el que la conciencia y las fuerzas mentales en general obtienen su debida representación como rasgos importantes de la cadena de control». ²⁹ Luego explicaba así su razonamiento: «En primer lugar, sostenemos que los fenómenos conscientes o mentales son propiedades dinámicas, emergentes y modeladoras (o configuradoras) del cerebro viviente en acción, un aspecto aceptado por muchos, incluidos algunos de los más rigurosos investigadores del cerebro. En segundo término, esta argumentación da otro importante paso más, y afirma que, en el cerebro, dichas propiedades modeladoras emergentes tienen potencia de control causal, como la tienen en otras partes del universo. Y ahí tenemos la respuesta al viejo enigma de la conciencia». Sperry partía de la base de que la experiencia consciente es una propiedad de la actividad cerebral irreductible (no se puede descomponer en sus partes), dinámica (cambia en respuesta a la actividad neuronal) y emergente (es más que la mera suma de los procesos que la producen), y afirmaba que no podía existir desligada del cerebro. Negando cualquier tipo de dualismo, subrayaba: «El término [*fuerzas mentales*] se ajusta al fenómeno de la experiencia subjetiva, pero aquí no implica la existencia de ninguna fuerza sobrenatural incorpórea independiente del mecanismo cerebral. Las fuerzas mentales, como aquí se conciben, se hallan ineludiblemente ligadas a la estructura cerebral y a su organización funcional». ³⁰ No hay, pues, fantasmas en el sistema.

A principios de la década de 1970 esta idea estaba obteniendo una cierta aceptación, y, de hecho, contribuyó al surgimiento de una creciente postura anticonductista. Las imágenes mentales, las ideas y los sentimientos internos volvían a estar sobre la mesa. Incluso podían tener un papel causal en las posibles explicaciones. Se había iniciado la «revolución cognitiva», que todavía continúa.

Los modernos filósofos prueban suerte

Mientras tanto, los filósofos debatían en torno a diversas teorías que incorporaban la visión materialista del cerebro. Tras su jubilación en 1975, Eccles dejó atrás el laboratorio y unió sus fuerzas al eminente filósofo Karl Popper. Ambos coincidían con Descartes en que, si la actividad mental es efectiva,³¹ es decir, si un pensamiento puede afectar a un estado cerebral, el cerebro tiene que estar abierto a influencias no físicas. Eccles trató de formular hipótesis verificables en ese sentido, pero no tuvo éxito, y finalmente se conformó con un modelo de interacción mente/cerebro sin ninguna evidencia experimental y sin ninguna hipótesis comprobable. Mientras que su versión de dualismo no generó precisamente muchos fans, hubo otra versión, de distinto carácter, cuya llama se reavivó aventada por las alas de un murciélago.

En 1974, el conocido filósofo Thomas Nagel, de la Universidad de Nueva York, publicó un llamativo artículo que llevaba por título «¿Cómo es ser un murciélago?», y que marcaría el comienzo de toda una línea de argumentación centrada en la cuestión de «cómo podemos explicar la experiencia de...» (por ejemplo, la rojez de lo rojo). Sosteniendo que la conciencia tiene un esencial carácter subjetivo (como argumentaba Franz Brentano), Nagel afirma que «un organismo tiene estados mentales conscientes si y solo si existe un *cómo es ser* ese organismo; un *cómo es ser* para dicho organismo».³² Ese cómo no tiene aquí un significado comparativo, tal que en la pregunta: «¿Cómo es patinar sobre hielo? ¿Es igual que patinar sobre ruedas?»; lejos de ello, alude a la sensación cualitativa subjetiva de la experiencia, es decir, de lo que se siente siendo ese sujeto, tal que en: «¿Cómo es para ti patinar sobre hielo?» (por ejemplo, ¿es emocionante?). Nagel lo denominó «carácter subjetivo de la experiencia»; también se le ha llamado «conciencia fenoménica», y, aunque él no lo menciona, se conoce asimismo como *qualia* (o «sensación cruda»)³³.

Según Nagel, hay algo que el sujeto de una experiencia «siente» al tener esa experiencia; para toda criatura hay un «cómo es ser» la especie que es y no otra, y el carácter subjetivo de un estado mental solo puede ser aprehendido

por ese sujeto concreto. Esta idea vino a ser «como» (aquí sí, en términos comparativos) servir un plato de pasta a la carbonara a un futbolista hambriento: fue devorada por los filósofos, que, según Peter Hacker — también filósofo—, llevaban tiempo buscando la salvación del «fiscalismo reduccionista o el funcionalismo sin alma». ³⁴ Para algunos, la vía de escape pasó a ser esta: la ciencia es objetiva, la conciencia es subjetiva; nunca llegarán a encontrarse, o, si lo hacen, lo harán gracias a nuevas leyes físicas o fundamentales aún no descritas (la postura actual de Nagel). ³⁵

No obstante, el filósofo Daniel Dennett se ha hecho célebre por cuestionar la pregunta de Nagel. Dennett afirma que en realidad Nagel no quiere saber cómo sería para él ser un murciélago, sino que lo que desea es saber objetivamente cómo es serlo subjetivamente: «No le bastaría con tener la experiencia de ponerse un “casco murcielagador” —un casco con electrodos que estimularían su cerebro para producir experiencias similares a las de un murciélago— y, en consecuencia, experimentar la “murcielaguez”. Al fin y al cabo, eso no dejaría de ser meramente cómo sería para Nagel ser un murciélago. ¿Qué le satisfaría entonces? No está seguro de que algo pueda hacerlo, y eso es lo que le preocupa. Teme que este concepto de “tener experiencia” esté fuera del alcance del objetivo». ³⁶

Es decir, fuera del alcance de la ciencia. Esta es la que muchos consideran la brecha infranqueable entre los ámbitos subjetivo y objetivo. El nuevo dualismo.

Dennett maneja este problema negándolo. Se lamenta de que uno de los obstáculos que se plantean a la hora de explicar la conciencia es que todos pensamos que somos expertos en conciencia, y tenemos creencias muy arraigadas al respecto, solo porque la hemos experimentado. Y se queja de que eso no les sucede, por ejemplo, a los investigadores de la visión: aunque la mayoría de nosotros podemos ver, no creemos que por ello seamos expertos en ese campo. Dennett afirma que la conciencia es el resultado de la chistera de un ilusionista: nuestra experiencia subjetiva es una ilusión, una muy creíble,

por la que siempre nos dejamos engañar aunque se nos haya explicado cómo se produce físicamente, como ocurre con algunas ilusiones ópticas que siguen engañándonos aun después de saber cómo funcionan.

El filósofo Owen Flanagan también discrepa de la idea de que exista una brecha insalvable, escribiendo: «Es fácil explicar por qué uno experimenta subjetivamente de manera única ciertos acontecimientos cerebrales: solo él está adecuadamente conectado a su propio sistema nervioso para tener sus propias experiencias». ³⁷ Esto parece razonable. Entonces, ¿dónde está el problema? Si bien la mayoría de los filósofos actuales aceptan que todo acontecimiento y experiencia mental es algún tipo de acontecimiento físico, muchos, sin embargo, se resisten a admitir la conclusión de que la esencia de un acontecimiento o experiencia mental puede captarse íntegramente mediante una descripción a nivel neuronal. Flanagan se limita a adoptar la postura de que no hay nada misterioso en el hecho de que los estados mentales conscientes posean un lado fenoménico: todo forma parte de la codificación.

De modo que nos deslizamos en la era moderna sin que nada se haya resuelto aún. Aunque la neurociencia había descubierto cómo funcionan los reflejos, cómo las neuronas se comunican entre sí, cómo se heredan los rasgos y muchas otras cosas, esta disciplina seguía sin tener el menor indicio de cómo el cerebro crea lo que se ha dado en llamar nuestra experiencia consciente fenoménica. Las ciencias que estudian la relación cerebro/mente no habían experimentado nada parecido a una revolución einsteniana, y, aunque la psicología cognitiva era capaz de explorar el interior de la caja negra, se aconsejaba a los jóvenes científicos que dejaran correr el tema de la conciencia.

Francis Crick a la ciencia moderna: sí se puede estudiar la conciencia

Dos décadas después de que George Miller aparcara la conciencia durante una

década o dos, habría de ser alguien como Francis Crick —un hombre siempre intrépido, de una inteligencia suprema, creativo y curioso— quien irrumpiría en escena para sacarla de nuevo a la palestra. Había dos incógnitas que intrigaban a Crick desde temprana edad: el origen de la vida y el enigma de la conciencia. Tras dedicar treinta años de su vida a la primera de ellas, estaba ansioso por afrontar la segunda. De modo que en 1976, a la tierna edad de sesenta años, cuando la mayoría de la gente solo piensa en la jubilación, hizo las maletas, dejó Cambridge y puso rumbo al Instituto Salk de San Diego para iniciar una segunda carrera científica en la neurociencia.

Casualmente, tuve ocasión de visitar el Salk poco después de su llegada, y allí me condujeron a su espectacular despacho con vistas al mar. Crick, que apenas empezaba a sumergirse en el campo de la neurociencia, estaba rodeado de otros investigadores de talento. Yo no tenía ni idea de cómo agregarme a la conversación, de modo que le pregunté: «¿Cómo se pueden concebir las escalas de tiempo prescritas por los procesos moleculares, y cómo se relacionan estas con las diferentes escalas de tiempo operativas en la actividad neuronal? Cada nivel tiene su historia; entonces ¿cómo se relacionan?». Pareció gustarle la pregunta, y unos meses después, alentado por aquella conversación, lo invité a un pequeño encuentro sobre la memoria que yo estaba organizando en la isla de Moorea. Aceptó al instante. Crick siempre se sentía molesto y se impacientaba cuando oía hablar del *statu quo* de cualquier tema. Le gustaban los buenos experimentos, pero siempre quería saber qué significaba una determinada observación en un contexto más amplio. En el encuentro no fue distinto: siempre dinamizaba las cosas. Parecía ser la persona que hacía falta para impulsar los estudios sobre la conciencia más allá de las posiciones clásicas ya consolidadas.

Crick empezó por aprender neuroanatomía y leer mucho sobre neurofisiología y psicofísica. En 1979, un par de años después de iniciada su aventura, le pidieron que escribiera un artículo para un número de la revista *Scientific American* dedicado a los últimos avances en la investigación sobre

el cerebro. Su tarea: «hacer algunos comentarios generales acerca de cómo veía el tema una persona relativamente ajena al mundillo». En el artículo, Crick señalaba que no le gustaba el modo en que los conductistas y funcionalistas abordaban el cerebro como una caja negra, ya que, al fin y al cabo, lo que estaba en cuestión era el propio funcionamiento interno de esa caja negra: «La dificultad del enfoque de la caja negra es que, a menos que esa caja sea intrínsecamente muy simple, pronto se llega a un estadio en el que varias teorías rivales explican todas ellas igualmente bien los resultados observados». ³⁸ Y nadie creía que la caja fuera intrínsecamente simple.

Crick también comentaba que los científicos que estudiaban el cerebro estaban demasiado aislados en sus subdisciplinas concretas. Tenían que ser científicamente menos provincianos y más cosmopolitas, participando en diálogos interdisciplinarios. Los psicólogos debían entender la estructura y función del cerebro, al igual que los anatomistas tenían que saber de psicología y fisiología. Obviamente, aquellas valoraciones generales de Crick no tenían en cuenta a docenas —por no decir centenares— de científicos cognitivos ³⁹ y de jóvenes neurocientíficos cognitivos en ciernes, como un servidor, ⁴⁰ que ya habían empezado a tratar de dilucidar con gran esfuerzo el problema de la conciencia. Pero fue él, gracias a su estatus único y especial, quien hizo ver a los investigadores de esta disciplina que el estudio de la base física de la conciencia era una tarea crucial.

Todos necesitaban un poco de neuropsicología y un poco de física y química; y Crick pensaba que el nuevo campo de la teoría de la comunicación resultaba prometedor como herramienta teórica, de modo que también había que hincarle el diente. Solo podría surgir una teoría general cuando pudieran explicarse todos los aspectos y niveles de las operaciones del cerebro y del comportamiento humano. Si solo estabas familiarizado con un aspecto, no tenías la menor posibilidad de dar con ningún tipo de explicación global.

Una de las sugerencias de Crick resultaba particularmente difícil: proponía que teníamos que cambiar la mentalidad de la gente con respecto a la

precisión de su propia introspección, puesto que «nuestra introspección nos engaña en todos los niveles». ⁴¹ Uno de los ejemplos que ponía de tal engaño era el punto ciego que todos tenemos en cada ojo. Crick también reprendía a los filósofos actuales —aunque cabe suponer que Dennett no estaba en la lista — por ignorar tales fenómenos:

No todo el mundo es consciente de que tiene un punto ciego, aunque es fácil de demostrar. Lo extraordinario es que no vemos un hueco en nuestro campo visual. La razón es, en parte, que no tenemos forma de detectar los bordes del hueco, y, en parte, que nuestro cerebro llena ese hueco con información visual que toma prestada de la zona inmediatamente circundante. Nuestra capacidad para engañarnos a nosotros mismos con respecto al funcionamiento de nuestro cerebro es casi ilimitada, sobre todo porque aquello de lo que podemos informar es solo una diminuta parte de lo que ocurre en nuestra cabeza. De ahí que una gran parte de la filosofía haya sido estéril durante más de dos mil años, y es probable que siga siéndolo mientras los filósofos no aprendan a entender el lenguaje del procesamiento de la información.

Eso no significa, sin embargo, que el estudio de nuestros procesos mentales mediante la introspección deba abandonarse por completo, como han pretendido hacer los conductistas. Hacer eso sería descartar uno de los atributos más significativos de lo que estamos tratando de estudiar. Pero el hecho es que la evidencia de la introspección nunca debe aceptarse al pie de la letra; debe explicarse en términos que no sean meramente los suyos propios. ⁴²

Y concluía Crick:

El sistema nervioso superior parece ser una combinación extremadamente ingeniosa de cableado de precisión y redes asociativas [...]. La red se divide en muchas subredes de menor tamaño, algunas de ellas en paralelo, otras organizadas de forma más secuencial. Además, la parcelación en subredes refleja tanto la estructura del mundo, externa e interna, como nuestra relación con él. ⁴³

Crick era en el fondo un teórico, con un especial talento para asimilar ideas y resultados experimentales de una amplia gama de disciplinas, mezclarlos y agitarlos como en una coctelera, y luego formular a partir de ahí nuevas teorías y nuevos experimentos. Sabía expresar claramente los profundos problemas que implicaba tratar de entender la experiencia consciente, y poseía la inestimable habilidad de hacer aquello que sugería William James: «El arte de

ser sabio es el arte de saber qué pasar por alto». Y, desde luego, él era un sabio.

Crick no tardó en formar equipo con el neurocientífico computacional Christof Koch, del Instituto de Tecnología de California, un hombre de gran inteligencia e ilimitada energía. Ambos decidieron abordar la cuestión de la conciencia partiendo del estudio del sistema visual de los mamíferos, un tema que ya había generado una gran cantidad de datos experimentales. Su propósito era aprender lo máximo posible sobre los primeros pasos del procesamiento de la información que recibe la corteza visual, con el objetivo último de descubrir lo que se conoce como «correlatos neuronales de la conciencia» (CNC): el conjunto mínimo de acontecimientos y mecanismos neuronales capaces de dar cuenta de un determinado percepto consciente.⁴⁴ Explica Koch: «*Tiene que haber una correspondencia explícita entre cualquier acontecimiento mental y sus correlatos neuronales. Otra forma de decirlo es que cualquier cambio en un estado subjetivo debe estar asociado a un cambio en un estado neuronal. Téngase en cuenta que lo contrario no tiene por qué ser necesariamente cierto: dos estados neuronales distintos del cerebro pueden resultar mentalmente indistinguibles*». ⁴⁵ Todo esto parece de lo más razonable y sencillo, dos cualidades que escasean en la investigación sobre la conciencia.

Para iniciar su investigación, Crick y Koch partieron de dos supuestos sobre la conciencia. El primero era que en cualquier momento dado algunos procesos neuronales activos se correlacionan con la conciencia, mientras que otros no, de modo que se preguntaron: ¿qué diferencia a unos de otros? El segundo supuesto, que calificaron de «provisional», era que «los diferentes aspectos de la conciencia (olor, dolor, visión, autopercepción, etc.) emplean todos ellos uno o quizá varios mecanismos comunes»;⁴⁶ así pues, si lograban entender el funcionamiento de uno de dichos aspectos, estarían en vías de entenderlos todos. Decidieron aparcar algunos debates para evitar perder tiempo discutiendo sobre ellos. Sorteando el *impasse* de la relación

mente/cuerpo, decidieron que, para examinar la conciencia científicamente, dado que todos tenían una idea aproximada de lo que se entendía por conciencia, no necesitaban definirla, evitando de ese modo los riesgos de una definición prematura.

Y puesto que optaban por adoptar una postura difusa en cuanto a su definición, Crick y Koch decidieron abordar de forma igualmente difusa su función y dejar de lado la cuestión de para qué sirve la conciencia. También optaron por presuponer que algunas especies de mamíferos superiores poseen determinados rasgos de conciencia, pero no necesariamente todos. Así, por ejemplo, se pueden tener ciertos rasgos clave de conciencia sin tener un lenguaje. Y aunque los animales inferiores puedan tener cierto grado de conciencia, de momento no querían lidiar con esa cuestión. Supusieron que la autopercepción era un aspecto autorreferencial de la conciencia, y también decidieron aparcarlo. Provisionalmente dejaron de lado asimismo la volición y la intencionalidad, junto con los estados hipnóticos y el sueño. Por último, prescindieron también de los *qualia* —que, como hemos visto, definen el carácter subjetivo de la experiencia, como, por ejemplo, la rojez de lo rojo—, considerando que, una vez descubrieran cómo veía el rojo una determinada persona, quizá se pudiera argumentar de forma plausible que todos nuestros rojos son de hecho iguales.

Crick y Koch reconocían que los CNC no iban a resolver el misterio de la conciencia. Lo que sí haría la identificación de los correlatos neuronales del procesamiento consciente (*versus* el no consciente) en pro del estudio empírico de la conciencia sería proporcionar una serie de restricciones a las posibles especificaciones de modelos neurobiológicamente plausibles. Confiaban en que dilucidar los CNC podría suponer un importante avance en la teoría de la conciencia similar al que había supuesto la estructura del ADN en la transmisión genética. Entender la arquitectura de la molécula de ADN y elaborar un modelo tridimensional de esta proporcionó una serie de pistas acerca de cómo se dividía y reproducía dicha molécula, lo que se

correlacionaba muy bien con la herencia mendeliana. Los primeros CNC definitivos que se descubrieran representarían un primer paso hacia una teoría de la conciencia, pero en sí mismos no explicarían los vínculos existentes entre esta última y la actividad neuronal. Eso es lo que hacen los modelos; y pronto empezarían a surgir unos cuantos.

Fue Crick quien abrió las compuertas: estudiar la conciencia volvía a ser interesante. En las dos décadas anteriores se había establecido una base, construida con montones de datos empíricos sobre los mecanismos cerebrales. De modo que se lanzó un ataque de naturaleza empírica, ayudado por un arsenal cada vez mayor de nuevos métodos, que ahora incluían la capacidad no solo de registrar la activación de neuronas individuales, sino también de controlarla (un objetivo anhelado por Crick que se había logrado gracias a la optogenética), además de diversas técnicas de obtención de imágenes del cerebro, pasando por la gran capacidad de procesamiento de datos que podían proporcionar los modernos ordenadores. Quienes decidieron seguir el consejo de Crick de que «la introspección nunca debe aceptarse al pie de la letra; debe explicarse en términos que no sean meramente los suyos propios», se encontrarían con que tendrían de sobra dónde elegir en lo que al procesamiento inconsciente del cerebro se refería, ya que empezaron a surgir como setas modelos neurobiológicos que trataban de explicar los vínculos entre la actividad neuronal y la conciencia utilizando elementos computacionales, informativos y neurodinámicos. Dichos modelos varían en función del nivel de abstracción que abordan —como veremos en el capítulo 5—, y, aunque algunos comparten rasgos comunes, ninguno de ellos explica todos los aspectos de la conciencia, y todavía ninguno ha llegado a generar un consenso generalizado.

En los próximos capítulos me propongo plantear una nueva idea y un nuevo marco para concebir el problema de la conciencia. Lo haré a la vez con humildad e inquietud, ya que tratar de contribuir al saber previamente generado por el panteón de grandes pensadores y científicos que aquí hemos

visto resulta, cuando menos, una tarea ingente. Sin embargo, hoy tenemos a nuestro alcance una vasta cantidad de nueva información que se acumula con rapidez, y que, con un poco de suerte, puede ofrecer una nueva perspectiva de cómo el cerebro realiza su magia. Las ideas de Descartes y otros pensadores del pasado de que la mente era algo que de algún modo «flotaba» por encima del cerebro, y las ideas de los nuevos mecanicistas de que la conciencia era una cosa monolítica generada por un único mecanismo o red, resultan simplemente erróneas. El argumento que planteo aquí es que la conciencia no es una «cosa»: *conciencia* es la palabra que utilizamos para describir la percepción subjetiva de una serie de instintos y/o recuerdos que se suceden en el tiempo en un organismo. Es, pues, un término «vicario» que representa el funcionamiento de un organismo viviente complejo. Y para entender cómo funcionan los organismos complejos necesitamos saber cómo se organizan las diversas partes del cerebro para proporcionar la experiencia consciente tal como la conocemos. Vamos a ello.

PARTE II

El sistema físico

4

Cómo fabricar cerebros módulo a módulo

Lo que es aquí, como ves, hace falta correr todo cuanto una pueda para permanecer en el mismo sitio. Si se quiere llegar a otra parte hay que correr por lo menos dos veces más rápido.

LA REINA ROJA EN LEWIS CARROLL, *A través del espejo*

Se diría que nuestro cerebro tiene una estructura tan inestable como la que parece exhibir a primera vista el Museo Guggenheim de Bilbao, diseñado por Frank Gehry; pero, como el propio Gehry suele decir, el museo no tiene «fugas»: funciona perfectamente. Gehry es un genio de la arquitectura que ha ampliado nuestra concepción de las estructuras físicas que realizan funciones útiles. También nuestro cerebro tiene una estructura física que realiza funciones útiles: hay un método en el aparente delirio de esa estructura de aspecto inestable de la que entendemos algunas partes, aunque todavía no la mayoría de ellas. A pesar de varios siglos de investigación, nadie entiende aún del todo cómo ese intrincado entramado de tejido biológico que reside en el interior de nuestra cabeza produce las experiencias de nuestra vida cotidiana. En todo momento se están produciendo en nuestro cerebro un número enorme de procesos eléctricos, químicos y hormonales; sin embargo, nosotros lo experimentamos como un todo unificado que marcha sobre ruedas. ¿Cómo es posible? De hecho, ¿cuál es la organización de nuestro cerebro que genera la unidad consciente?

Todo tiene una estructura subyacente; los físicos llevan esta verdad hasta el

nivel cuántico (que analizaremos en el capítulo 7). Desmontamos constantemente las cosas para ver qué las hace funcionar. Las cosas están compuestas de partes, y lo mismo ocurre con el cuerpo y el cerebro. En ese sentido se podría decir que estamos hechos de módulos, esto es, de componentes que interactúan para producir el funcionamiento de la entidad completa que estamos examinando. Necesitamos conocer las partes, y no solo el modo como estas encajan, sino también cómo interactúan.

Hay pocas dudas de que, de un modo u otro, las diversas partes del cerebro trabajan conjuntamente para producir nuestros estados mentales y nuestros comportamientos. A primera vista, parece lógico pensar que el cerebro funciona como una unidad global para producir una única experiencia consciente. El premio Nobel Charles Sherrington, en un texto escrito a principios del siglo XX, describía ya el cerebro como un «telar encantado»,^{1 2} sugiriendo que el sistema nervioso funciona de manera coherente para crear la mente mística. Sin embargo, los neurólogos de la época le habrían sugerido que se diera una vuelta por unos cuantos hospitales: sus clínicas estaban llenas de pacientes cuyas lesiones cerebrales contaban una historia muy distinta.

Paradójicamente, aunque cada uno de nosotros «se siente» una sola entidad indivisa (un hecho que parece proporcionar la evidencia intuitiva que respalda la metáfora del telar de Sherrington), un considerable número de evidencias sugieren que el cerebro «no» funciona de una manera holística; lejos de ello, nuestra conciencia indivisa está producida en realidad por miles de unidades de procesamiento relativamente independientes o, por decirlo de manera más sencilla, por «módulos». Los módulos son redes especializadas y frecuentemente localizadas de neuronas que cumplen una función específica.

El neurocientífico, físico y filósofo Donald MacKay comentaba en cierta ocasión que resulta más fácil entender cómo funciona algo cuando no funciona correctamente. Por su trabajo en el ámbito de las ciencias físicas, sabía que los ingenieros podían determinar más deprisa cómo funcionaba algo,

pongamos por caso, un televisor, cuando la imagen parpadeaba que cuando funcionaba sin problemas.³ De manera similar, estudiar los cerebros enfermos nos ayuda a comprender mejor cómo funcionan los que no lo están.

Las evidencias más convincentes en favor de una arquitectura modular del cerebro surgen del estudio de pacientes que han sufrido una lesión cerebral. Cuando el daño se produce en determinadas áreas localizadas del cerebro, algunas capacidades cognitivas se verán alteradas debido a que la red de neuronas responsables de dichas capacidades dejan de funcionar, mientras que otras permanecen intactas y siguen funcionando tan campantes. Lo que resulta más intrigante de los pacientes con alteraciones cerebrales es que, independientemente de cuál sea su anormalidad, todos parecen conservar la conciencia. Si la experiencia consciente dependiera del buen funcionamiento de todo el cerebro en su conjunto, eso no debería ocurrir. Dado que este hecho —el carácter generalizado de la modularidad— ocupa un papel central en mi tesis, es importante entender hasta qué punto es realmente modular el cerebro.

Cerebros que funcionan aunque falten módulos

Cojamos un lóbulo, cualquier lóbulo del cerebro, y consideremos el caso de las personas que han sufrido un derrame cerebral. Las que sufren una lesión del lóbulo parietal derecho, por ejemplo, suelen padecer un síndrome denominado heminegligencia espacial. Dependiendo del tamaño y la ubicación de la lesión, los pacientes con heminegligencia pueden comportarse como si parte o la totalidad del lado izquierdo de su mundo no existiera, a veces incluyendo el lado izquierdo de su propio cuerpo. Esto puede llevarles a hacer cosas como no comerse la parte izquierda de su plato, no afeitarse o no ponerse maquillaje en el lado izquierdo de la cara, no dibujar el lado izquierdo de un reloj, no leer las páginas de la izquierda de un libro y no reconocer nada ni a nadie que esté en la mitad izquierda de la habitación. Algunos incluso niegan que su brazo y su pierna izquierda sean suyos, y no los

utilizan cuando tratan de levantarse de la cama aunque de hecho no los tengan paralizados. Algunos pacientes incluso llegan a prescindir del lado izquierdo del espacio en su imaginación y en sus recuerdos.⁴ El hecho de que estas deficiencias varíen en función del tamaño y la ubicación de la lesión sugiere que el daño que perturba determinados circuitos neuronales concretos se traduce en alteraciones de diferentes procesos integrantes. El mapeo de la neuroanatomía funcional de estas lesiones ha proporcionado firmes evidencias en favor de esta hipótesis.⁵

Ahora bien —y esta es la parte interesante—: aunque puede producirse heminegligencia cuando existe una pérdida real de sensación o de sistemas motores, también puede darse otra versión de este trastorno cuando todos los sistemas sensoriales y motores se hallan en buen estado de funcionamiento, un síndrome conocido como «extinción». En este caso, cada mitad del cerebro parece funcionar bien por sí sola, pero empieza a fallar cuando se requiere que funcione al mismo tiempo que la otra mitad. Pese a ello, la información del lado ignorado puede utilizarse a nivel inconsciente,⁶ lo que significa que dicha información está ahí, pero el paciente no es consciente de ello. He aquí cómo funciona. Si a un paciente con heminegligencia izquierda se le muestran estímulos visuales tanto en su campo visual derecho como en el izquierdo, este comunica que solo ve el estímulo de la parte derecha. En cambio, si se le muestra solo el estímulo visual izquierdo, que él recibe en la misma parte exacta de la retina que antes, dicho estímulo se percibe normalmente. En otras palabras, si no hay competencia por parte del lado normal, la información procedente del lado alterado se percibe correctamente y aflora en la percepción consciente. Lo más extraño de todo es que estos pacientes niegan que algo vaya mal: no son conscientes de la pérdida de esos circuitos ni de los problemas resultantes.

Parece, pues, que su yo autobiográfico debe de derivarse únicamente de aquello de lo que son conscientes. Y esto último depende de dos condicionantes. En primer lugar, no son conscientes de los circuitos que no

funcionan: es como si dichos circuitos nunca hubieran existido y la conciencia de lo que hacían hubiera desaparecido con ellos. En segundo término, se está produciendo algún tipo de procesamiento competitivo: el procesamiento de determinados circuitos aflora a la conciencia, mientras que el de otros no. En suma, la experiencia consciente parece estar vinculada a un procesamiento que es excesivamente local, que produce una capacidad específica; y ese procesamiento puede verse competitivamente superado por el procesamiento de otros módulos y no aflorar nunca a la conciencia, lo cual tiene unas implicaciones asombrosas.

Mientras que algunos pacientes no son conscientes de determinadas partes de su cuerpo que realmente están ahí, mi trastorno clínico favorito es el conocido como fenómeno «del tercer hombre»,⁷ en el que una persona siente la presencia de otra que en realidad no está ahí. Conocido en el ámbito científico como «sensación de presencia» (SdP), es la impresión de que hay alguien más que está presente en una determinada ubicación espacial, a menudo justo a nuestra espalda. La sensación es tan fuerte que la persona que la experimenta vuelve constantemente la cabeza para mirar o incluso para ofrecer comida a la presencia. No se trata de la misma sensación que uno tiene cuando camina por un callejón oscuro y se asusta imaginando que alguien le sigue. Esta presencia aparece de manera inesperada, y, de hecho, es un fenómeno común entre los alpinistas y otras personas que sufren un intenso agotamiento físico en condiciones extremas.

En su libro *La montaña desnuda*,⁸ Reinhold Messner, considerado por muchos el mejor alpinista de todos los tiempos (fue el primero en escalar solo el Everest y, dicho sea de paso, nunca utiliza oxígeno suplementario), describe lo que le sucedió en 1970 cuando realizaba su primera gran ascensión en el Himalaya, la del Nanga Parbat, en compañía de su hermano Günther: «De repente había un tercer escalador a mi lado. Descendía con nosotros, manteniendo una distancia regular un poco a mi derecha y a unos pasos de mí, justo fuera de mi campo de visión. Yo no podía ver la figura y seguí

manteniendo la concentración, pero estaba seguro de que había alguien allí. Podía sentir su presencia; no necesitaba más pruebas». Sin embargo, no hace falta ser un alpinista exhausto para experimentar tal presencia: casi la mitad de las personas que se han quedado viudas han sentido la presencia de su cónyuge fallecido.⁹ Para algunos, este fenómeno es el punto de partida de historias sobre apariciones, fantasmas e intervenciones divinas.

No es así, afirma el neurólogo y neurofisiólogo suizo Olaf Blanke, que se encontró con este fenómeno de manera inesperada tras excitar con estimulación eléctrica la corteza parietal temporal del cerebro de un paciente mientras intentaba localizar el foco de un ataque.¹⁰ Blanke también estudió a un grupo de pacientes que sufrían SdP, y descubrió que existen lesiones del área frontoparietal específicamente asociadas a este fenómeno, y que dichas lesiones se hallan siempre en el lado opuesto del cuerpo a aquel en el que se percibe la presencia.¹¹ Esta ubicación le sugirió la idea de que el fenómeno puede deberse a alteraciones del procesamiento sensoriomotor y la integración multisensorial. Aunque somos conscientes de nuestra ubicación en el espacio, no lo somos de la gran cantidad de procesos (visión, sonido, tacto, propiocepción, motricidad, etc.) que, cuando están normalmente integrados, nos ubican correctamente en él. Si existe un trastorno de ese procesamiento, pueden producirse errores, y nuestro cerebro puede malinterpretar nuestra ubicación. Blanke y sus colegas han descubierto que uno de dichos errores se manifiesta en forma de SdP. Hace poco lograron inducir hábilmente SdP en sujetos sanos alterando su procesamiento sensorial con ayuda de un brazo robótico.¹²

Cuando hacemos un movimiento, esperamos que su consecuencia se produzca en un momento y lugar del espacio concretos. Si te rascas la espalda, esperas experimentar una sensación en ella de manera simultánea. Cuando la sensación coincide espacial y temporalmente con lo que esperas, tu cerebro interpreta esa sensación como autogenerada. Si no existe coincidencia, si las señales son espacial y temporalmente incompatibles con el hecho de tocarte a

ti mismo, determinas que proceden de algún otro agente. Ahora imaginemos que tenemos los ojos vendados y los brazos extendidos hacia delante, con la yema del dedo introducida en una ranura en forma de dedal de un robot, que, a su vez, envía señales a un brazo robótico situado a nuestra espalda. El movimiento del dedo controla el del brazo robótico, que nos acaricia la espalda en función de dicho movimiento. En algunas de las pruebas, nuestro dedo encuentra una resistencia que se corresponde con la fuerza con la que está presionando, mientras que en otras la resistencia es aleatoria y no guarda una correlación apropiada con lo que hacemos. Si la sensación que percibimos en la espalda es sincrónica con nuestro movimiento, a pesar de tener los brazos extendidos nuestro cerebro crea una ilusión: sentiremos como si nuestro cuerpo se hubiera desplazado hacia delante y realmente nos estuviéramos tocando la espalda con el dedo. Si, por el contrario, la sensación táctil no es sincrónica, si se produce ni que sea una fracción de segundo más tarde, nuestro cerebro elabora algo distinto: la percepción de nuestra propia ubicación se desplaza en dirección opuesta, alejándose de la yema del dedo, y no sentiremos que seamos nosotros mismos quienes nos tocamos la espalda. Si, además, tampoco notamos resistencia en la yema del dedo mientras controlamos el brazo, este contacto asincrónico nos producirá la sensación de que quien nos toca la espalda es otra persona situada detrás de nosotros. Blanke, utilizando estimulaciones corporales muy bien controladas, demostró que los conflictos sensoriomotores (es decir, las señales que resultan espacial y temporalmente incompatibles con el autocontacto físico) bastan para inducir SdP en voluntarios sanos; esos conflictos se producían mediante la manipulación de diferentes redes neuronales localizadas, esto es, de diferentes módulos.

Si el cerebro funcionara realmente como un «telar encantado» conjunto, eliminar partes de este o estimular el procesamiento erróneo en algunos circuitos o bien provocaría la interrupción completa del sistema o bien causaría una disfunción en todos los ámbitos cognitivos. Por el contrario,

muchas personas pueden llevar una vida relativamente normal aunque les falte o tengan dañada una parte del cerebro. Cuando alguien tiene una lesión en determinadas áreas cerebrales concretas, casi siempre parece haber un deterioro en algunos dominios cognitivos, pero no en todos. Así, por ejemplo, un dominio cognitivo bien desarrollado en los humanos es el lenguaje. En la mayoría de las personas, el centro del lenguaje está ubicado en el hemisferio izquierdo, y dentro de este se hallan dos regiones muy distintas: el área de Broca y el área de Wernicke.

El área de Broca contribuye a la producción del habla, mientras que la de Wernicke se ocupa de la comprensión o interpretación del lenguaje hablado y escrito, y ayuda a organizar nuestras palabras y oraciones de manera inteligible. Más concretamente, el área de Broca está involucrada en la articulación de las palabras, coordinando los músculos de los labios, la boca y la lengua para pronunciarlas con precisión, mientras que la de Wernicke organiza las palabras en un orden comprensible incluso antes de que hablemos. Las personas con el área de Broca dañada tienen dificultades para hablar: les cuesta articular las palabras y hablan en ráfagas, pero las palabras que logran pronunciar brotan en un orden comprensible (por ejemplo, «cerebro... modu... lar»), aunque sus frases puedan no ser gramaticalmente correctas. Estos pacientes son conscientes de sus errores, y no tardan en sentirse frustrados. En cambio, las personas que tienen alguna lesión en el área de Wernicke sufren sobre todo un trastorno de comprensión: su pronunciación es normal y su habla exhibe una gramática correcta, pero lo que dicen no tiene sentido. Estas evidencias nos revelan que cada una de esas dos áreas ejerce una función distinta y específica; si una de ellas está dañada, ya no puede realizar correctamente dicha función, lo cual demuestra de manera fehaciente que en el cerebro existe una modularidad hiperespecífica.

¿Por qué surgió esta modularidad cerebral? En cierta ocasión oí al presidente de Coca-Cola describir la lógica de la organización corporativa de la empresa. Cuando esta última alcanzó un cierto tamaño, los ejecutivos se

dieron cuenta de que tener una planta central que fabricara íntegramente su producto y luego lo enviara a todo el mundo resultaba descabellado, ineficiente y costoso. El transporte, los costes de embalaje, los gastos de viaje de los directivos para reunirse en la «sede central» de la empresa, etc., etc. Carecía de sentido. Era obvio que tenían que dividir el mundo en diferentes zonas, construir plantas en cada una de ellas y luego distribuir su producto a escala local. Adiós a la planificación centralizada; bienvenido el control local. Lo mismo ocurre con el cerebro: resulta más económico y eficiente.

El desarrollo de un cerebro de mayor tamaño

Históricamente siempre se ha supuesto que los animales con cerebros más grandes de lo que cabría esperar dado su tamaño corporal tenían mayor inteligencia y habilidades. Se creía que los humanos tenían cerebros excesivamente grandes en proporción a su cuerpo, y eso explicaba nuestras peculiares habilidades e inteligencia. Sin embargo, esta teoría siempre ha planteado un problema: los neandertales tenían un cerebro mayor que el nuestro, pero, en cambio, no lograron superar la competencia del *Homo sapiens* cuando este último apareció en escena. A partir de mi propia investigación se deriva otro problema igualmente espinoso: tras practicar una operación quirúrgica de «cerebro escindido», o callosotomía, el hemisferio cerebral izquierdo, ahora aislado (es decir, la mitad del cerebro), resulta ser casi tan inteligente como lo era el cerebro entero intacto. Por lo tanto, mayor no significa necesariamente mejor. ¿Qué está ocurriendo aquí?

Suzana Herculano-Houzel y sus colegas, armados con una nueva técnica para contar el número de células neuronales y no neuronales de los cerebros humanos, se dedicaron a comparar dicho número con el de otras especies, y descubrieron que los rumores sobre nuestros grandes cerebros eran de lo más exagerados. El cerebro humano no resulta desproporcionado en lo que a tamaño se refiere, sino que está en perfecta proporción con la escala de los

primates. Aunque el cerebro humano es mucho mayor y tiene muchas más neuronas, la *ratio* entre el número de neuronas y el tamaño del cerebro es la misma que en los chimpancés.¹³ Otro hallazgo estimulante fue descubrir que otra cifra citada con frecuencia (aunque barajada sin mencionar jamás ninguna referencia que la sustentara), la *ratio* de 10:1 que supuestamente existía entre células gliales y neuronas en el cerebro humano, resultaba absolutamente exagerada. De hecho, el cerebro humano no tiene más de un 50 % de células gliales, igual que los cerebros de otros primates. Rompiendo otro mito más, Herculano-Houzel sugiere asimismo que esa sobreestimación de la *ratio* entre células gliales y neuronas podría ser el origen de la falsa idea de que solo utilizamos un 10 % de nuestro cerebro.¹⁴

Sin embargo, en comparación con los cerebros de otros mamíferos, el cerebro humano presenta dos ventajas. En primer lugar, está construido según unas reglas de escalado que permiten economizar y ahorrar espacio y que coinciden con las de otros primates; además, entre todos esos cerebros de primates tan económicamente contruidos es el mayor, y, por lo tanto, el que contiene el mayor número de neuronas. Sin embargo, no se puede utilizar a tontas y a locas el tamaño del cerebro como equivalente al número de neuronas cuando se comparan los primates con otras especies. Por ejemplo, en los roedores, si comparamos ratones y ratas, el cerebro de la rata es mayor, pero no solo porque tenga más neuronas: en el caso de la rata, al aumentar el número de neuronas, también lo hizo el tamaño de estas; por lo tanto, una sola neurona de rata ocupa más volumen que una sola neurona de ratón (es algo parecido a la diferencia de tamaño entre los espaguetis y el cabello de ángel). En cambio en los primates, cuando se comparan los monos con los humanos, aunque el número de neuronas aumente, el tamaño de estas sigue siendo el mismo. El resultado es que un cerebro de primate más grande representa un mayor incremento neto en el número de neuronas por volumen que un cerebro de roedor más grande. Si cogiéramos un cerebro de rata y aumentáramos su volumen hasta hacerlo igual al de uno humano, la rata tendría solo una séptima

parte del número de neuronas del cerebro humano debido simplemente al hecho de que cada una de ellas ocuparía más espacio. Aumentar el tamaño del cerebro es un asunto complicado, y se diría que los diferentes órdenes de mamíferos (*Primata, Rodentia...*) han seguido reglas distintas en lo relativo a su escalado.

Todo esto nos lleva de vuelta a los módulos. Si, a medida que fue aumentando el número de neuronas del cerebro humano, cada una de ellas se hubiera conectado con todas las demás, el número de axones (los «cables» de cada neurona) se habría incrementado de manera exponencial. Nuestros cerebros serían gigantescos; de hecho, tendrían veinte kilómetros de diámetro,¹⁵ y requerirían tanta energía que aunque nos forzaran a alimentarnos como una oca de engorde no seríamos capaces de hacerlo funcionar.¹⁶ En realidad nuestro cerebro representa aproximadamente el 2 % de todo nuestro peso corporal y absorbe alrededor del 20 % de nuestra energía. Los cerebros usan tanta energía porque son potentes sistemas eléctricos que están constantemente activos, como un acondicionador de aire en pleno agosto. Otro problema del supuesto anterior sería que los axones serían tan largos que las velocidades de procesamiento bajarían en picado.

El neurocientífico Georg Striedter, que estudia cómo y por qué se produjeron diferencias en la evolución del cerebro en diversas especies, sugiere que hay ciertas leyes que gobiernan la conectividad cuando los cerebros aumentan de tamaño.¹⁷ En primer lugar, el número de neuronas a las que se conecta una neurona media no varía con dicho incremento; por el contrario, el número absoluto de conexiones neuronales sigue siendo el mismo, con el resultado de que el aumento de tamaño del cerebro sería más manejable en términos de requerimientos de energía y espacio. Eso implica, no obstante, que al agrandarse el cerebro se produce una disminución de la conectividad global. Y una menor conectividad implica un procesamiento más independiente.

La segunda ley es que las longitudes de las conexiones se minimizan. Esto

se traduce en que la mayoría de las neuronas están conectadas con las neuronas vecinas. Las conexiones cortas consumen menos energía, menos espacio y menos tiempo en la transmisión de señales, produciendo una comunicación eficiente entre esas neuronas localizadas. Por lo tanto, cuando los cerebros se agrandan, se produce una reorganización de sus conexiones y cambia su arquitectura estructural. La arquitectura estructural resultante es una basada en grupos de neuronas localizadas bien conectadas entre sí, denominados «comunidades».

Este tipo de organización permite que esos grupos separados se especialicen de forma independiente en la realización de una determinada función: así nace un módulo. Aunque la mayoría de las neuronas de un módulo tienen conexiones intramodulares, un pequeño número de ellas establecen también conexiones cortas con neuronas de módulos vecinos, lo que permite la formación de un circuito neuronal. Se forma un circuito neuronal cuando un módulo recibe información, la modifica y la transmite a otro módulo para una ulterior modificación. Por lo tanto, aunque la mayoría de los módulos apenas establecen conexiones con otros módulos, esta interconexión permite que los módulos vecinos formen a su vez grupos que posibilitan un procesamiento más complejo. Aprenderemos más al respecto en el próximo capítulo, cuando analicemos la arquitectura estratificada.

Algunos módulos están organizados jerárquicamente, de modo que se hallan formados por submódulos, que a su vez se hallan formados por subsubmódulos.¹⁸ Sin embargo, la existencia de varios módulos que funcionan de manera independiente crea la necesidad de una comunicación y coordinación eficientes entre ellos; lo cual nos lleva a la tercera ley de la interconexión neuronal, que sería: no todas las conexiones se minimizan, sino que se mantienen algunas conexiones largas que sirven como «atajos» entre sitios distantes.

La arquitectura global que producen estas leyes de interconexión es la que se conoce como arquitectura «de mundo pequeño». Este tipo de arquitectura es

notorio por su capacidad de albergar un alto grado de modularidad, y, sin embargo, requerir pocos pasos para interconectar dos módulos cualesquiera seleccionados de manera aleatoria. La arquitectura de mundo pequeño está presente en muchos sistemas complejos, como las redes sociales o determinadas redes eléctricas. Numerosos estudios han corroborado la hipótesis de que el cerebro está organizado en grupos o módulos de regiones funcionalmente interconectadas.¹⁹

Ventajas de un cerebro modular

Al observar este diseño, podemos constatar que hay muchas razones por las que un cerebro modular es superior a uno que funcione de manera global. Para empezar, un cerebro modular reduce los costos energéticos. Como está dividido en unidades, solo hace falta que determinadas regiones de un módulo dado estén activas para realizar cada tarea concreta. Si tuviéramos que utilizar todo nuestro cerebro para cada acción, el consumo eléctrico nos reventaría el cráneo. Es lo mismo que ocurre con el aire acondicionado en pleno agosto: si por las noches solo lo mantienes encendido en el dormitorio, sale mucho más barato que enfriar la casa entera; pero, aunque ahorre energía gracias a la modularidad, ¿de verdad el cerebro es energéticamente tan eficiente si resulta que tenemos que dedicar una quinta parte de nuestra dieta a alimentarlo?

Lo cierto es que el cerebro es bastante eficiente a pesar de ser un devorador de energía. Las neuronas transmiten impulsos eléctricos a través de los axones y las dendritas, que forman los «cables» del cerebro. Aunque este cableado neuronal funciona de manera muy distinta al de los modernos dispositivos eléctricos, la idea básica es la misma: una corriente eléctrica transfiere información de un lugar a otro, y eso requiere energía. Cuanto más lejos se desplaza una corriente eléctrica, más energía consume, y cuanto más grueso es el axón, más resistencia encuentra y, por lo tanto, más energía requiere para superarla. Al trabajar en módulos locales, el cerebro ahorra

energía operando en distancias cortas, con cables delgados y con tiempos breves de transmisión de la información que viaja entre esos módulos. Asimismo, dada la dinámica de los sistemas neuronales, si se minimiza la longitud del cable y, por ende, los retrasos en la transmisión, cabe predecir una proporción óptima de cableado (la proporción de materia gris formada por los axones) del 60 %; y, de hecho, muchas estructuras cerebrales están compuestas de sistemas de cableado que se acercan a este valor óptimo.²⁰ Si, por el contrario, el cerebro funcionara como una unidad global, cada región cerebral tendría una cantidad de cableado aproximadamente igual para la comunicación a corta y a larga distancia; y no olvidemos que las distancias más largas implican más «cable» y, por lo tanto, mayor «coste». En cambio, un cerebro modular reduce costes al mantener el cableado en una ratio menor de 3 a 5 (la mencionada proporción de cable del 60 %), lo que limita el número de transmisiones eléctricas de larga distancia. En términos generales, pues, el cerebro parece maximizar la eficiencia energética al operar en módulos.

Los cerebros modulares también son funcionalmente eficientes, puesto que varios módulos pueden procesar información especializada de manera simultánea. Es mucho más fácil caminar, hablar y mascar chicle a la vez si existen muchos sistemas modulares que funcionan de manera independiente, en lugar de un solo sistema intentando coordinar todas las acciones. Además, si el cerebro actuara como una sola unidad, tendría que ser una especie de «aprendiz de todo, maestro de nada», para realizar de forma adecuada todas nuestras tareas cotidianas; obviamente, resulta más conveniente disponer de módulos «maestros» especializados en tareas específicas. La especialización es una característica omnipresente en los sistemas complejos. Así, por ejemplo, una economía prospera cuando los mejores agricultores se dedican a cultivar, los mejores educadores a educar y los mejores gestores a gestionar. Un mal gerente puede hundir un negocio; un mal agricultor puede arruinarse, y un mal maestro... Bueno, todos hemos tenido que sufrir al menos a uno de ellos y ya sabemos las consecuencias. Cuando las personas se dedican a tareas

concretas y se centran en ellas sin tener que preocuparse por todas las demás labores necesarias para mantener una economía en funcionamiento se convierten en expertas; y los expertos son productores más eficientes. Cuando los expertos trabajan simultáneamente se genera un mayor rendimiento económico del que habría si todo el mundo intentara hacer un poquito de todo. Por lo tanto, parece razonable pensar que nuestros cerebros evolucionaron de una forma modular para procesar de forma eficiente múltiples tipos de información a la vez.

Y lo que quizá resulta aún más importante: un cerebro modular también permite una adaptación o evolución más rápida del sistema en un entorno cambiante; dado que un módulo puede modificarse o duplicarse independientemente del resto, no hay riesgo de alterar o perder otros módulos que ya estén bien adaptados en ese proceso. En consecuencia, la ulterior evolución de una parte no amenaza los aspectos del sistema que ya estén funcionando bien.

Incluso si eliminamos la evolución de la ecuación, la modularidad cerebral resulta útil para la adquisición de nuevas habilidades. Los investigadores han descubierto que durante el aprendizaje de una habilidad motora la arquitectura de determinadas redes se modifica.²¹ Aunque llegar a dominar muchas habilidades requiere un tiempo considerable, el hecho es que podemos aprenderlas a través de la práctica. Si cada vez que adquirimos una nueva habilidad el cerebro entero tuviera que cambiar su forma de funcionamiento, perderíamos la práctica adquirida en las que ya poseíamos. Así pues, las ventajas de la modularidad cerebral son que ahorra energía cuando los recursos son escasos, nos permite el procesamiento cognitivo especializado en paralelo cuando el tiempo es limitado, hace que resulte más fácil alterar la funcionalidad cuando surgen nuevas presiones de supervivencia y nos permite aprender toda una serie de nuevas habilidades. Si uno se para a pensar en ello, ¿cómo podría el cerebro haberse organizado de otro modo?

Cómo hacerse modular

Los cerebros humanos no son ni los únicos cerebros ni los únicos sistemas biológicos modulares. El cerebro de los gusanos, el de las moscas y el de los gatos son también modulares, como lo son las redes vasculares, las redes de interacciones proteína-proteína, las redes de regulación genética, las redes metabólicas y hasta las redes sociales humanas.²² ¿Cómo evolucionó esta modularidad? ¿Qué presiones selectivas producen un sistema modular? Esta era la pregunta que desconcertaba a un trío de informáticos que, tras reflexionar sobre ello, decidieron poner a prueba la hipótesis de Striedter de que la modularidad es un subproducto de la presión para minimizar los costes de conexión.²³

Los costes de construcción de una red incluyen los costes de fabricación de las conexiones y su mantenimiento, la energía que requiere la transmisión por ellas y el coste de los retardos de propagación de la señal. Cuanto más largas sean las conexiones y mayor sea su número, más caro resultará construir y mantener la red.²⁴ Asimismo, añadir más conexiones o aumentar la longitud de una vía de transmisión podría retrasar los tiempos de respuesta crítica, lo que no resulta conveniente para la supervivencia en un entorno competitivo cuando un depredador empieza a salivar al verte, te enseña los colmillos y afila las garras.

Los informáticos Jeff Clune, Jean-Baptiste Mouret y Hod Lipson hicieron lo que suelen hacer los informáticos: diseñaron simulaciones por ordenador.²⁵ Para ello utilizaron redes ya bien estudiadas que admitían entradas sensoriales y generaban ciertas señales de salida; lo que determinaban estas últimas era lo bien que funcionaba la red cuando tenía que hacer frente a problemas del entorno. Simularon 25.000 generaciones de evolución, programando una presión selectiva directa, o bien para maximizar únicamente el rendimiento, o bien para maximizar el rendimiento y «al mismo tiempo» minimizar los costes de conexión. Y *voilà!* Una vez que se añadió la minimización de los costes de cableado, tanto en entornos cambiantes como

invariables, de inmediato empezaron a surgir módulos, mientras que cuando no existía ese requisito estos no aparecían. Y cuando los tres científicos observaron cómo eran las redes de mayor rendimiento así desarrolladas, dichas redes resultaron ser modulares. En este grupo descubrieron asimismo que, cuanto menores eran los costes, mayor era la modularidad resultante. Además, esas redes también evolucionaban mucho más deprisa —en un número bastante reducido de generaciones—, ya fuera en entornos estables o cambiantes. Estos experimentos de simulación proporcionan firmes evidencias de que las presiones selectivas para maximizar el rendimiento de la red y minimizar los costes de conexión producen redes que son significativamente más modulares y más capaces de evolucionar.

Así pues, sabemos ya que los sistemas modulares tienen muchas ventajas. Pero ¿cómo se las arreglan? ¿Cómo miles de módulos localizados e independientes consiguen funcionar juntos para coordinar nuestros pensamientos y conductas y, en última instancia, producir nuestra experiencia consciente?

Conexiones modulares

Aunque los módulos se hallan extremadamente intraconectados para poder realizar funciones especializadas, hemos aprendido que a la vez están poco interconectados con otros módulos. Ahora bien, algunas comunicaciones entre módulos son vitales para coordinar comportamientos complejos. Así, por ejemplo, las áreas de Broca y de Wernicke tienen sus propias funciones especializadas relacionadas con el lenguaje, pero también deben conversar entre sí. El área de Wernicke necesita organizar los fonemas y las palabras en oraciones coherentes para que el área de Broca guíe nuestros labios, boca y lengua a fin de producir la secuencia de sonidos correcta. De modo que estas dos áreas especializadas en el lenguaje se encuentran densamente interconectadas a través del denominado fascículo arqueado, un conjunto de

fibras nerviosas que discurre entre ellas como si fuera una autopista. Nuestro cerebro minimiza la necesidad de estas redes de comunicación largas y costosas reduciendo las conexiones entre módulos que contribuyen a diferentes tipos de funciones cognitivas. Por ejemplo, no hace falta activar las áreas de Broca y de Wernicke cuando uno está oliendo una rosa, a menos que se lance a declamar un soneto alabando su belleza o una diatriba contra los cultivadores por producir híbridos favoreciendo la forma por encima de la fragancia. Los módulos cerebrales se comunican entre sí, pero existe un número desproporcionadamente mayor de conexiones entre aquellos módulos que realizan procesos cognitivos interrelacionados, mientras que entre los módulos que participan en procesos distintos el número de conexiones es mucho menor.

Cerebro animal y humano: ¿cuál es la diferencia?

Aun cuando emplean métodos y técnicas de análisis de datos sorprendentemente distintos, la mayoría de los estudios demuestran que los módulos, tanto en redes cerebrales estructurales como funcionales, se dan en todas las especies y comparten muchas de las mismas propiedades.²⁶ Vale la pena tomarse un momento para entender la diferencia entre una red estructural y una funcional. Por «estructura» se entiende aquí simplemente la anatomía física de una red: cuántas neuronas hay, cómo están organizadas, su forma, etc. Una red funcional es aquella que realiza una determinada función; puede tener que ver con la capacidad de hablar una lengua o con la de entenderla. Es importante destacar que la estructura de una red no revela su función, ni tampoco sucede a la inversa. Puede dar pistas sobre ella, pero eso es todo. Por ejemplo, podemos observar un árbol y ver su estructura, pero eso no nos dice nada sobre la función de las hojas. Varios estudios realizados con diversas especies de animales, desde invertebrados hasta mamíferos, han revelado que sus módulos neuronales también se hallan altamente intraconectados y espacialmente cerca unos de otros a fin de reducir el

consumo de energía. Curiosamente, la red neuronal del nematodo transparente *Caenorhabditis elegans* (un organismo que posee unos cientos de neuronas que han sido ampliamente estudiadas) también funciona a base de módulos a pesar de ser una de las criaturas más diminutas de todas las que poseen un sistema neuronal.²⁷ En diferentes especies, la modularidad se revela eficiente y necesaria para que los organismos funcionen y evolucionen de manera eficaz en un entorno competitivo.

Parece lógico suponer que, si los cerebros modulares están presentes en animales y humanos, deben de compartir aspectos cognitivos similares, incluida la conciencia. Desafortunadamente, por mucho que pudiera gustarle a Thomas Nagel, la tecnología actual no nos permite comprender realmente cómo los diferentes organismos experimentan el mundo; a menudo incluso nos resulta difícil entender nuestra propia percepción de este. Lo máximo que podemos hacer para poder entender de forma empírica la experiencia de otros, tanto animales como personas, es utilizar indicadores conductuales y de la actividad cerebral.

No resulta sorprendente que asociemos la experiencia consciente con nuestras complejas habilidades cognitivas humanas; y de ahí pasamos a concluir que, para que un animal sea consciente, ha de tener el mismo tipo de habilidades. De hecho, proyectamos libremente la capacidad de experimentar la conciencia en toda clase de cosas, desde títeres hasta robots y, en mi caso, incluso un Plymouth cupé de 1949.

Uno de los métodos que han empleado los investigadores para buscar pistas sobre posibles estadios de conciencia tempranos en otros animales es el de tratar de detectar signos de uso de herramientas, ya que se considera que utilizar herramientas es un comportamiento que señala la presencia de una cognición compleja. Pero resulta que hay signos de ello en todo el reino animal. Así, por ejemplo, los córvidos (la familia *Corvidae*: cuervos, cornejas, arrendajos, urracas, grajos, cascanueces) desarrollan herramientas para obtener alimentos de lugares difíciles de alcanzar de forma similar a

como los chimpancés fabrican y usan herramientas.²⁸ Los cuervos de la ciudad japonesa de Sendai utilizan automóviles para descascarillar frutos secos: los depositan en los pasos de peatones, y no solo esperan a que pase un coche y los aplaste, sino que esperan a que el semáforo se ponga rojo para recogerlos. Los cuervos de Nueva Caledonia son auténticos prodigios: fabrican dos tipos de herramientas, que utilizan de diferentes maneras para realizar distintos trabajos. Cuando van en busca de comida las llevan consigo del mismo modo que un pescador llevaría su caña. También resuelven los denominados problemas de «metaherramientas», en los que tienen que utilizar una herramienta para obtener una segunda, necesaria a su vez para obtener un determinado alimento.²⁹ Los cuervos de diferentes zonas utilizan distintos diseños de herramientas, lo que sugiere que entre ellos se da una variación y transmisión culturales,³⁰ aunque las habilidades básicas relacionadas con el uso de herramientas de palo pueden desarrollarlas incluso cuervos criados en cautividad sin ningún aprendizaje social.³¹ Si bien esto probablemente significa que los cuervos son conscientes en el sentido de que están vivos, alerta y experimentan el momento, ¿indica también que son conscientes de sus habilidades? Seguramente tienen algunos módulos especializados que no están presentes en otras aves; pero ¿los hace eso conscientes de sí mismos? Lo cierto es que los numerosos estudios realizados sobre su comportamiento, habilidades y aprendizaje no se arriesgan a abordar esta cuestión. ¿Y qué hay de los chimpancés?

Hace ya tiempo que se ha observado el uso de herramientas en chimpancés en su hábitat natural, principalmente palos para recoger hormigas y miel, y hojas para recoger agua. Los chimpancés de diferentes ubicaciones geográficas utilizan asimismo distintas herramientas para propósitos diversos, lo que también aquí sugiere que existe una variación cultural y una transmisión social del uso de herramientas. Sin embargo, una vez que un chimpancé ha aprendido un comportamiento asociado al uso de una herramienta, este se convierte en un hábito, y en general los animales de esta especie no se

actualizan al uso de una nueva técnica mejorada aunque unos pocos miembros de su grupo descubran y utilicen una.³² Por otra parte, se ha observado que los chimpancés que han vivido con humanos son capaces de resolver rompecabezas y de hallar soluciones a problemas complejos. Por ejemplo, los chimpancés que veían un plátano colgando del techo fuera de su alcance apilaban cajas unas encima de otras a fin de crear una escalera improvisada que les permitiera alcanzarlo.³³ Si bien la lista de trucos de los chimpancés es larga y deslumbrante, ¿les convierte eso en seres conscientes en el mismo sentido en que lo son los humanos? Probablemente esta sea una pregunta mal planteada; quizá la pregunta correcta debería ser: «¿Nuestra experiencia consciente tiene contenidos similares a la de un chimpancé?».

Ignorando el significado de todas estas investigaciones con animales, muchos estudios se han dedicado a comparar el pensamiento de los chimpancés con el de los niños pequeños. En una sencilla prueba consistente en señalar objetos ocultos, en la que un objeto deseado se coloca fuera de la vista y el experimentador señala dónde está, los chimpancés se quedan perplejos, mientras que los bebés humanos la superan ya a los catorce meses de edad.³⁴ Al mismo tiempo, si un chimpancé o un niño observa un determinado comportamiento, ambos son capaces de imitarlo aunque nunca antes hayan realizado esa clase de acciones; sin embargo, mientras que los niños imitan todas las acciones que se les muestran a fin de obtener una recompensa, incluso las más superfluas, los chimpancés solo imitan las necesarias para ello. Se ha sugerido que eso indica que los niños son imitadores compulsivos, mientras que los chimpancés imitan para alcanzar un objetivo. Si un chimpancé no recibe una recompensa (o un castigo), generalmente el comportamiento aprendido no se repite; en cambio, los bebés imitan los comportamientos independientemente de que impliquen una recompensa o un castigo, lo que parece sugerir que los bebés humanos tienen tendencia a aprender nuevos comportamientos por el mero impulso de aprender.³⁵ De ser cierto, esto marcaría una enorme diferencia entre los

humanos y el resto del reino animal. En cualquier caso, parece que los chimpancés tienen más actividad que los córvidos. ¿Es posible que su hardware craneal añadido permita una experiencia consciente o simplemente cambia su contenido?

Los humanos tienen una capacidad de aprender y resolver problemas abstractos que supera la de otros animales. El hombre ha inventado tecnologías que son más sofisticadas y proporcionan mucha más utilidad que cualquier herramienta que haya creado ningún animal. Ingenieros y científicos han construido ordenadores, aviones, rascacielos, cohetes que nos llevan a la Luna... todo lo habido y por haber. Sin embargo, solo necesitamos que una pequeña parte de la población sea inventiva. Mediante la imitación y el aprendizaje, los objetos y descubrimientos útiles se propagan como el fuego entre la población y pasan a formar parte de nuestra vida cotidiana. Como ha señalado el eminente psicólogo David Premack, los humanos disponen de un reducido «grupo selecto» capaz de desarrollar grandes tecnologías, como el dominio del fuego, la rueda, la agricultura, la electricidad, los teléfonos móviles, Internet o las patatas rellenas. Ninguna otra especie viviente cuenta con ningún miembro que pueda realizar tales hazañas.³⁶ ¿Es esta capacidad adicional de aprender, resolver problemas e inventar la que nos permite ser conscientes? Tiene que haber un hardware, esto es, unos módulos especiales, que nos permitan tales capacidades. ¿Son dichos módulos la clave de nuestro entendimiento?

Personalmente, toda esa línea de pensamiento que sostiene que hay una especie de poción mágica que produce la conciencia humana me parece descaminada. Ver todas las cosas maravillosas que puede hacer un chimpancé hace que nuestra mente reaccione y, entonces, le conferimos un estatus especial; lo admitimos en nuestro club de la conciencia y estamos encantados de hacerlo. Pero eso mismo llevó a la persona que inicialmente descubrió y describió la vida mental de los chimpancés a plantearse la siguiente pregunta:

¿qué piensan ellos de todo esto? Nosotros tenemos una teoría sobre los chimpancés, pero ¿tienen ellos una teoría sobre nosotros?

Premack y su alumno Guy Woodruff fueron los primeros en intentar comprobar si otros animales tienen lo que se denomina una «teoría de la mente».³⁷ Eso significa que un individuo es capaz de atribuir estados mentales —como, por ejemplo, propósito, intención, conocimiento, creencias, dudas, fingimiento, agrado, etc.— tanto a sí mismo como a los demás. Premack y Woodruff, que fueron quienes acuñaron la expresión, la denominaron «teoría» porque tales estados no son directamente observables en otros individuos, sino que se infieren. Los humanos damos por supuesto que nuestro prójimo tiene mente y que sus estados mentales impulsan sus acciones. Casi cuarenta años después de que se propusiera la idea, el polvo que levantó aún no se ha asentado, pero parece que, si bien algunos animales poseen cierto grado de teoría de la mente, ninguno alcanza el nivel de los humanos. Josep Call, Michael Tomasello y sus colegas han dedicado muchos años a acotar esta cuestión. Los chimpancés entienden hasta cierto punto los objetivos e intenciones de otros individuos, así como sus percepciones y conocimientos; pero, pese a los numerosos intentos de demostrar lo contrario, parece ser que estos animales no son capaces de entender que otros puedan albergar falsas creencias,³⁸ una prueba que los niños de dos años y medio superan sin problemas.³⁹ Hace poco, no obstante, Call y Tomasello, en colaboración con Christopher Krupenye, han hallado evidencias que sugieren que tres especies de grandes simios son capaces de entender de manera implícita que otros individuos pueden albergar falsas creencias, pero todavía no se ha demostrado que tomen decisiones conductuales explícitas basadas en dicha comprensión.⁴⁰ Está por ver en qué medida la teoría de la mente de los simios puede acercarse a la de los humanos.

Recientemente los perros han empezado a compartir también el centro de atención en las investigaciones sobre el cociente intelectual de los animales en relación con la sociabilidad. *Chaser*, el famoso border collie del profesor de

psicología ya jubilado John Pilley, conoce más de mil palabras, entiende la sintaxis y es capaz de hacer inferencias sobre el posible significado de las palabras nuevas.⁴¹ Si, por ejemplo, le dicen que traiga el *dax* (una palabra que nunca había oído antes), examinará su voluminoso montón de juguetes y traerá el que no había visto hasta entonces. Los perros también pueden hacer inferencias sobre comida escondida y otros objetos ocultos basándose en pistas sociales, como que un humano señale hacia algún sitio (cosa que los chimpancés no hacen). Michael Tomasello sugiere que esto implica que son capaces de entender dos niveles de intención: «qué» y «por qué». En primer lugar, el perro debe entender que la señal pretende que preste atención a aquello hacia lo que se señala; y, en segundo término, también debe deducir por qué se supone que tiene que hacerlo: ¿está ofreciéndole esa persona información útil acerca de dónde se encuentra algo, o bien desea el objeto para sí?⁴² Aunque los chimpancés suelen prestar atención al gesto de señalar, no entienden que allí hay comida escondida; es decir, no parecen ser capaces de deducir el segundo nivel de intención, el porqué. En los últimos doce años, la impresionante capacidad de los perros de utilizar las pistas de comunicación proporcionadas por humanos ha suscitado el interés de algunos investigadores que estudian la teoría de la mente, y, si bien existen algunos primeros indicios que sugieren que los perros parecen tenerla en cierta medida,⁴³ se necesita investigar mucho más en esta materia.

Aunque los amantes de los perros están encantados con esos hallazgos, hay que recordar que estos animales no muestran ninguna flexibilidad especial en los ámbitos no sociales: son esclavos de determinados intereses concretos; quiero decir, de determinadas capacidades concretas. Cuando se les proporcionan pistas no sociales, como esconder comida bajo un tablón inclinado en lugar de hacerlo bajo uno plano, son incapaces de utilizarlas para resolver el problema (que, en cambio, es fácil de resolver para un chimpancé); ni entienden que deberían optar preferentemente por tirar de la cuerda que tiene comida atada en lugar de tirar de la que no la tiene (de nuevo, algo que

un chimpancé es capaz de identificar de inmediato).⁴⁴ Las diferentes capacidades cognitivas de los perros sugieren que poseen módulos específicos pero distintos, que evolucionaron en respuesta a presiones ambientales diversas. Los contenidos de su experiencia consciente son distintos de los nuestros y diferentes de los de los chimpancés, aunque, sin duda, hay algunos comunes.

En términos generales, parece que intentar establecer un requisito cognitivo previo para la experiencia consciente es un ejercicio fútil. Coger un poquito de aquí y un poquito de allá no sirve para determinar qué tiene que hacer el cerebro para engendrar una experiencia consciente. El cerebro no va a revelar fácilmente ese truco, si es que de hecho es un truco. Recuérdese que no experimentamos de manera consciente el punto ciego de nuestro campo visual a pesar de que está ahí: nuestro sistema visual realiza un truco de conciencia. Pero para la mayoría de las personas la experiencia consciente no es ningún truco: es algo muy real, algo que gestiona una parte o sistema del cerebro, y ya se ha levantado la veda para descubrirlo. Dado que los humanos poseen un procesamiento cognitivo avanzado que permite el desarrollo y la utilización de nuevas tecnologías y la formulación de inferencias sobre las creencias y los deseos de los demás, la pregunta es: ¿acaso el cerebro humano tiene algo que el de los animales no tiene?

Un reciente estudio comparativo analizó el volumen de neuropilo de diferentes áreas del cerebro humano y el del chimpancé.⁴⁵ El neuropilo es el conjunto de todas las regiones cerebrales que están formadas por conexiones: una mezcla de axones, dendritas, sinapsis y demás. La corteza prefrontal humana —el área del cerebro responsable de la toma de decisiones, la resolución de problemas, la atribución de estados mentales y la planificación temporal— tiene un porcentaje de neuropilo mayor del que se encuentra en el cerebro del chimpancé, mientras que las dendritas de esta región tienen un número mayor de espinas con las que conectarse a otras neuronas que las de otras áreas del cerebro. Este hallazgo anatómico sugiere que los patrones de

conectividad de las neuronas prefrontales pueden contribuir a lo que sea que diferencie nuestro cerebro. Curiosamente, los córvidos tienen un prosencéfalo relativamente mayor que el de la mayoría de las otras aves, sobre todo aquellas áreas que se cree que son análogas a la corteza prefrontal de los mamíferos.⁴⁶ Sin embargo —como veremos—, aunque esta concepción puede explicar las mayores capacidades de las que gozamos, no nos conducirá al objetivo de entender cómo se activa la conciencia. Reincidir en el supuesto de que hay un ingrediente secreto o una región especial del cerebro que nos proporciona la experiencia consciente no es un buen punto de partida.

¿Dónde está la conciencia?

Tenemos que cambiar de marcha. Hemos de deshacernos de la idea del ingrediente secreto, de que existe un lugar o un algo especial. Debemos pensar más bien en un conjunto de módulos en gran parte independientes y en cómo su organización da lugar a nuestra omnipresente sensación de experiencia consciente. Como científicos cognitivos, nos hemos entusiasmado demasiado con la idea de que la conciencia es un fenómeno independiente del resto de nuestros procesos psicológicos. En lugar de ello, deberíamos concebirla como un aspecto intrínseco de muchas de nuestras funciones cognitivas: si perdemos una determinada función, perdemos la conciencia que la acompaña, pero no perdemos completamente la conciencia.

Una primera pista de que la conciencia no está ligada a una red neuronal concreta proviene de mis propios estudios en pacientes operados de callosotomía. Aunque hay más conexiones neuronales dentro de cada hemisferio cerebral de las que existen entre ambos, el número de interconexiones entre los dos hemisferios sigue siendo enorme. Aun así, cortar dichas conexiones influye poco en la sensación de experiencia consciente. Es decir, que el hemisferio izquierdo sigue hablando y pensando como si nada hubiera ocurrido a pesar de que ya no tiene acceso a la mitad de la corteza

cerebral. Y lo que es más importante: desconectar los dos hemisferios crea instantáneamente un segundo sistema consciente, también independiente. Ahora el hemisferio derecho funciona sin preocuparse para nada del izquierdo, con sus propias capacidades, deseos, metas, ideas y sentimientos. Una sola red, dividida por la mitad, pasa a convertirse en dos sistemas conscientes. ¿Cómo pensar, entonces, que la conciencia surge de una red específica? Necesitamos una nueva idea para dar cuenta de este hecho.

Consideremos, también, cómo es la experiencia consciente para el paciente de cerebro escindido que se despierta de la operación, así como la circunstancia de que ahora cada uno de los dos hemisferios ignora la información procedente del campo visual del otro. El hemisferio izquierdo no ve el lado izquierdo del espacio, mientras que el derecho no ve el lado derecho. Sin embargo, el hemisferio izquierdo —responsable del habla del paciente— no se queja de ninguna pérdida de visión. De hecho, el paciente nos dirá que no nota absolutamente ninguna diferencia después de la cirugía. ¿Cómo puede ocurrir tal cosa cuando la mitad del campo visual ha desaparecido? Al igual que en el caso de los pacientes con heminegligencia espacial, el hemisferio izquierdo «parlante» nunca se queja de haber perdido la mitad de su campo visual, ya que los módulos responsables de informar de dicha pérdida se han quedado en el hemisferio derecho y ya no pueden comunicarse con el izquierdo, que por su parte ni los echa en falta ni sabe siquiera que alguna vez estuvieron allí. El recuerdo de haber tenido ese campo visual también ha desaparecido del hemisferio izquierdo. Ahora la experiencia consciente del campo visual derecho solo la disfruta el hemisferio derecho, y desaparece por completo de la experiencia del hemisferio izquierdo. ¿Qué nos dice eso sobre la conciencia?

Tras haber descartado la idea de un único módulo «consciente», podemos empezar a centrarnos en lo que realmente es la conciencia. Sabemos que las lesiones cerebrales locales pueden producir diversas discapacidades cognitivas específicas. Sin embargo, los pacientes que las sufren siguen siendo

conscientes del mundo que los rodea. El paciente con una heminegligencia espacial grave no es consciente de la mitad izquierda del espacio, pero sigue siéndolo de la derecha.

¿Y si la experiencia consciente está gestionada por cada módulo? Se pierde un módulo debido a una lesión o a un derrame cerebral, y la conciencia que acompaña a dicho módulo se va con él. Recuerde que los pacientes con heminegligencia no son conscientes de la mitad del espacio porque el módulo que procesa esa información ya no funciona. O lo que ocurre cuando los módulos responsables de ubicarnos en el espacio no se integran de forma adecuada: la experiencia consciente se ve profundamente alterada, y uno termina sintiendo que hay alguien más a su espalda. O el caso de las personas que padecen la enfermedad de Urbach-Wiethe, que provoca el deterioro de la amígdala: esas personas dejan de sentir miedo. A pesar de haber sido objeto de extensos estudios durante más de veinte años, una de estas pacientes todavía sigue sin hacerse una idea clara de la deficiencia que padece, y con frecuencia se ve en medio de situaciones de peligro.⁴⁷ Al parecer, el hecho de no tener la experiencia consciente del miedo le impide evitar dichas situaciones.

Esta idea de que la conciencia es una propiedad de cada módulo individual —y no una red única asociada a una especie determinada— podría explicar los diferentes tipos de conciencia que existen en las diversas especies. Los animales no son zombis inconscientes, pero aquello de lo que cada uno de ellos es consciente difiere en función de los módulos que tiene y de cómo están interconectados dichos módulos. Los humanos tenemos una rica experiencia consciente gracias a los numerosos tipos de módulos que poseemos. De hecho, es muy posible que tengamos módulos integradores altamente desarrollados, que nos permitirían combinar información procedente de varios módulos en pensamientos abstractos. Es difícil descifrar cómo surge la conciencia en los humanos, pero concebir esta última como un aspecto del funcionamiento de múltiples módulos puede guiarnos hacia la respuesta.

Aun así, si la conciencia es un aspecto de varios dominios cognitivos distintos, ¿cómo es que las personas con el cuerpo calloso cortado siguen experimentando el mundo como una sola entidad en lugar de una serie de retazos aleatoriamente presentados en cada momento dado? Para entender esto, podemos relacionar el procesamiento del cerebro con una especie de competición. Los módulos experimentan constantes variaciones en cuanto a la cantidad de actividad eléctrica que poseen, con el resultado de que su aportación a nuestra experiencia consciente también varía. La idea sería que el módulo más «activo» gana la competición de la conciencia, y su procesamiento se convierte en la experiencia vital, el «estado» del individuo en un determinado momento del tiempo. Imagine que está en una playa observando el vuelo de un ave exótica. En ese momento la emoción visual, la visión del pájaro y sus plumas llenas de colores, ha ganado la contienda de la experiencia consciente. En el momento siguiente la que gana la competición es la llamada de otro pájaro; y el siguiente, un sentimiento de curiosidad que nos hace volver la cabeza para localizar la fuente del sonido. De repente pasa a tener prioridad un fuerte dolor en el pie, lo que de inmediato nos lleva a mirar hacia abajo para ver a un cangrejo agarrado a uno de nuestros dedos. En cada momento del tiempo, nuestra única experiencia consciente es el aspecto cognitivo que ocupa el lugar más destacado en nuestro entorno exterior o interior; es «la rueda que chirría». Pero los diversos procesos que han competido entre sí son obra de diferentes módulos. ¿Cómo funciona todo esto?

Mi hipótesis es que lo que denominamos «conciencia» es una percepción que constituye un telón de fondo de, o vinculado a, un determinado acontecimiento o instinto mental actual. Esto se entiende mejor si utilizamos como ejemplo una arquitectura común que en ingeniería se conoce como «estratificación», y que permite a toda clase de sistemas complejos funcionar de manera eficiente e integrada, desde los átomos y las moléculas hasta las capacidades cognitivas y perceptuales, pasando por las células y los circuitos. Si el cerebro realmente está integrado por diferentes «capas» (en el sentido

mencionado), la información procedente de un micronivel puede integrarse en niveles de creciente complejidad hasta que cada unidad modular en sí misma produzca la conciencia. Una arquitectura estratificada (es decir, formada por diversas capas) permite que surjan nuevos niveles de funcionamiento a partir de capas de niveles inferiores que por sí solas no podrían crear esa experiencia de «nivel superior». Es hora de aprender más sobre la estratificación y las maravillas que aporta a la comprensión de la arquitectura cerebral. Estamos en vías de entender que la conciencia no es una «cosa», sino que es el resultado de un proceso incardinado en una arquitectura, del mismo modo que una democracia tampoco es una cosa, sino el resultado de un proceso.

Empezamos a entender la arquitectura cerebral

En arquitectura hay muchas cosas que permanecen ocultas para el ojo poco avezado.

FRANK GEHRY

Imagínese que es usted el joven científico en ciernes que sus padres creen que es. Por Navidad le regalan un viejo despertador y le dicen: «De acuerdo, señor sabelotodo, desmóntalo y vuelve a montarlo, y mientras lo haces explícanos cómo funciona». Sería bastante fácil. En el despertador solo hay un pequeño número de piezas —ruedas, engranajes, muelles—, que trabajan juntas en una determinada arquitectura para producir una función, y sabemos cuál es esa función. Pero resultaría mucho más difícil si no tuviéramos ni idea de cuál era la función y solo tuvieras las piezas en la mano.

Para aquellos de nosotros que estudiamos el cerebro humano, el problema radica en determinar cómo 89.000 millones de neuronas se conectan entre sí para sustentar nuestro material cognitivo. Los investigadores diseccionan los cerebros, los tiñen, los punzan, los cartografían y les ponen escuchas. Se han recopilado minuciosamente enormes cantidades de datos, se ha estudiado a pacientes con lesiones y se han examinado las hazañas mentales de personas excepcionales en busca de la desconcertante magia subyacente que estamos intentando entender. El congreso anual de la Sociedad de Neurociencia congrega nada menos que a veintiséis mil neurocientíficos para intercambiar datos e ideas, y aun así esta disciplina todavía está buscando un marco en el que situar toda esa información. ¿Por qué este resulta tan escurridizo? ¿Qué se

nos está pasando por alto? Debe de haber otra dimensión del problema que todavía no se ha logrado captar. A mediados del siglo XX, el biólogo teórico Robert Rosen le sugirió a su hija un posible dilema: «El cuerpo humano cambia por completo la materia de la que está hecho aproximadamente cada ocho semanas, a través del metabolismo, la replicación y la reparación. Sin embargo, tú sigues siendo tú, con todos tus recuerdos, tu personalidad... Si la ciencia insiste en perseguir partículas, las seguirá por todo el organismo, pero pasará absolutamente por alto este último». ¹

La observación de Rosen sugiere que la organización tiene que ser independiente de las partículas materiales que conforman un sistema viviente. De hecho, los componentes estructurales y la función del cerebro son solo una parte de la historia. Se necesita un tercer componente —a menudo ignorado— que relacione la estructura de un sistema con su función. Lo que falta saber es cómo se organizan las partes, los efectos de las interacciones entre estas y las relaciones con el tiempo y el entorno. Un profesor de Rosen, Nicolas Rashevsky —físico teórico y matemático de la Universidad de Chicago—, denominó «biología relacional» al estudio de dicho componente. Esta concepción ha sido asumida también por los investigadores en ingeniería electrotécnica y biología de sistemas, pero en general los biólogos moleculares y neurocientíficos la desconocen o la ignoran a pesar de que han transcurrido cincuenta años desde que Rosen formulara su advertencia.

Mi propia experiencia con esta forma alternativa de concebir la organización del cerebro se inició de la mano de John Doyle, profesor de control y sistemas dinámicos, ingeniería electrotécnica y bioingeniería en el Instituto de Tecnología de California. La primera lección del doctor Doyle fue que aprender sobre las partes no puede llevarte más allá de estas. Una escuela tiene lugares para leer y comer, para lavarte las manos y almacenar material. Lo mismo ocurre en una casa. Sin embargo, una escuela no es igual que una casa: ambas sirven a funciones distintas y el flujo de personas que pasa por ellas es completamente distinto. Una importante diferencia radica en la

organización de las partes, en su arquitectura. El erudito húngaro-británico Michael Polanyi explicaba que una «máquina en su conjunto funciona bajo el control de dos principios distintos. El superior es el principio del diseño de la máquina, y este se sirve del inferior, que consiste en los procesos físico-químicos en los que se basa la máquina». ² El diseño de la máquina constriñe de algún modo la naturaleza con el fin de aprovecharla para realizar una determinada tarea. Así, por ejemplo, una cafetera está hecha de piezas adecuadamente diseñadas que encajan de tal manera que produce una taza de café. Polanyi sostiene que esas restricciones imponen lo que se denominan «condiciones de frontera» a las leyes de la física y la química, y explica que los organismos son sistemas que comparten esta característica con las máquinas: «El organismo, al igual que una máquina, se revela como un sistema que funciona de acuerdo con dos principios distintos: su estructura actúa como una condición de frontera que se sirve de los procesos físico-químicos mediante los cuales sus órganos realizan sus funciones. Por lo tanto, este sistema puede calificarse como un sistema sometido a un doble control». ³ El diseño al que se refiere Polanyi es la arquitectura del organismo, y es la clave para entender el complejo mente/cerebro. Se trata de una idea crucial.

La arquitectura de lo complejo

Doyle es un maestro a la hora de conceptualizar cómo los sistemas complejos —cosas como un Boeing 777 o nuestro cerebro, ambos integrados por numerosas piezas que interactúan entre sí— pueden funcionar de manera eficiente, rápida y segura en lugar de explotar, estrellarse o detenerse entre chirridos. No debe sorprendernos que el término *complejidad* —al igual que ocurre con el de *conciencia*— no goce de una definición universalmente aceptada. Pero, para nuestro propósito actual, podemos centrarnos en tres dimensiones concretas de la complejidad en un sistema dado. Un sistema es complejo si tiene un gran número o un número diverso de: a) componentes; b)

interconexiones e interacciones, y *c*) comportamientos resultantes, algunos de ellos predecibles y otros no tanto. Actualmente los sistemas artificiales han empezado a tener niveles de complejidad casi biológicos.⁴ Por ejemplo, Doyle calcula que un Boeing 777 cuenta con 150.000 módulos de subsistemas distintos, organizados en sistemas de control y redes complejos, incluyendo unos mil procesadores que hacen volar el avión. Aunque, obviamente, los componentes de los sistemas tecnológicos avanzados y los sistemas biológicos altamente evolucionados son distintos, comparten ciertas similitudes en su arquitectura organizativa.⁵

Normalmente, la palabra *arquitectura* evoca el arte y la ciencia de diseñar edificios y otras estructuras como puentes y autopistas, su estilo de diseño (barroco, modernista...) y su método de construcción (tapial, vidrio y acero...). Quizá nos venga también a la mente algún pensamiento sobre, por ejemplo, Brunelleschi o Palladio. Pero por «arquitectura» se entiende también la estructura compleja de algo. Ese algo no es necesariamente un edificio, y puede o no ser un objeto físico: podría ser la estructura de mando de un gobierno o las vías de comunicación de Internet o las redes neuronales de nuestro cerebro. La arquitectura, en su nivel fundamental, tiene que ver con el diseño dentro de los límites de ciertas restricciones. Estas últimas son las condiciones de frontera de Michael Polanyi: límites impuestos por un poder restrictivo absoluto.⁶ En el caso de un edificio, eso significa trabajar dentro de los límites que imponen los materiales utilizados (paja, barro, madera, ladrillo, piedra, acero), el lugar en el que se construirá (regiones conocidas por sufrir frecuentes incendios, inundaciones, terremotos o huracanes; terrenos llanos o escarpados; zonas tropicales o tundra), la función del edificio (vivienda, sala de conciertos o gasolinera), y, obviamente, los deseos de los propietarios (la causa final de Aristóteles), entre otras cosas. En el caso de nuestro cerebro y sistema nervioso, las restricciones arquitectónicas incluyen los costes energéticos, el tamaño y las velocidades de procesamiento.

Los sistemas biológicos y tecnológicos complejos comparten una

arquitectura altamente organizada: es decir, que los componentes del sistema se organizan de una manera específica que permite su funcionalidad y/o robustez. Por poner un sencillo ejemplo: las fibras de algodón de una prenda de ropa presentan una arquitectura extremadamente organizada que hace que la tela resulte funcional para vestir.⁷ Es robusta frente al desgaste que normalmente sufre la ropa. En cambio, mezclar de forma arbitraria esas mismas fibras de algodón da como resultado papel, que no resiste el mismo grado de desgaste. La similitud de la arquitectura de los sistemas complejos y organizados sugiere que todos ellos comparten unos requisitos universales. Están diseñados para ser «eficientes, con capacidad de adaptación, susceptibles de evolucionar y robustos».⁸

Lo robusto, lo complejo y lo frágil

Entender que los animales, grandes y pequeños, no son muy distintos en cuanto a diseño de nuestro BMW o camioneta permite hacerse una idea más clara de cómo se las apaña el tejido biológico. Doyle y su colega David Alderson sostienen que la complejidad de los sistemas altamente organizados no es accidental, sino el resultado de estrategias de diseño (ya sean artificiales o evolutivas) que potencian la robustez o, como habría dicho Darwin, la aptitud.

Doyle y Alderson definen la robustez del siguiente modo: «Una [propiedad] de un [sistema] es robusta si es [invariante] con respecto a un [conjunto de perturbaciones]»,⁹ donde los corchetes indican que es necesario especificar cada uno de esos términos. Tomemos nuevamente como ejemplo un sistema complejo fácil de entender para ilustrar este concepto: la ropa. Supongamos que estamos haciendo las maletas para ir de viaje porque queremos ver la aurora boreal. Esta se produce en invierno en las zonas árticas y nosotros no queremos pasar frío. Llevar ropa [sistema] de plumón [propiedad] en nuestro gélido viaje puede representar una opción robusta, puesto que nos mantendrá calientes cuando bajen las temperaturas. Pero si hay una lluvia torrencial [una

perturbación que no se había especificado] y se nos moja la chaqueta, el plumón no nos mantendrá calientes. Aunque el plumón es (relativamente) invariante frente a las temperaturas frías, no lo es frente a la humedad; es decir, que resulta robusto para determinadas condiciones, pero frágil para otras. Si, en cambio, elegimos «me da un aspecto más estilizado» como [propiedad] y «aumento de peso» como [perturbación], terminaremos optando por ropa más ligera y exhibiendo un esbelta figura en medio de una ventisca (una opción robusta en cuanto al aumento de peso), pero nuestra ropa no será invariante a la perturbación del descenso de la temperatura.¹⁰

Cada rasgo que incrementa la robustez de un sistema protege a dicho sistema de ciertos desafíos internos o externos. Al mismo tiempo, cada nuevo paso hacia la robustez también incrementa la complejidad. Por desgracia, ninguna función adicional es robusta para todas las contingencias. Cada función añadirá su propio talón de Aquiles al sistema, una nueva vulnerabilidad a un nuevo reto imprevisto. De modo que, una vez descubierta esta nueva fragilidad, habrá que añadir un nuevo rasgo para combatirla, pero, al mismo tiempo, el siguiente rasgo comportará una nueva fragilidad frente a la que habrá que protegerse. Cada nueva protección aumenta la complejidad, lo que a su vez exige una complejidad aún mayor.

Estos pros y contras relacionados con cada propiedad (o rasgo) hacen inevitablemente que el comportamiento del sistema se vuelva más robusto frente a determinadas perturbaciones, pero frágil frente a otras. Estos rasgos «robustos pero frágiles» son característicos de los sistemas complejos altamente evolucionados. El concepto de robustez, siempre acompañado del de fragilidad, se encuentra en todas partes. Uno de mis ejemplos biológicos favoritos procede de los estudios sobre el desarrollo cerebral.

Es obvio que las conexiones neuronales son importantes para la función cerebral normal. Las neuronas de un determinado lugar del cerebro deben estar conectadas con las de otro lugar para posibilitar la coordinación última de actividad que genera el comportamiento. La evolución parece garantizarlo

de manera bastante sólida asegurándose de que exista una enorme sobreproducción de neuronas durante el desarrollo. Así, en lugar de que la estructura A envíe estrictamente la cantidad justa de neuronas a la estructura B, envía muchas más para garantizar la robustez. La madre naturaleza ha descubierto un modo de deshacerse de las neuronas sobrantes a través de un proceso conocido como «poda». Con la contribución apropiada del entorno, las neuronas innecesarias mueren, y al final del período de desarrollo hay un número aceptable conectando las dos estructuras. Pero, como no podía ser de otro modo, aquí se produce también la fragilidad. Con frecuencia la poda es exagerada; de hecho, se han acumulado una serie de evidencias de que los errores producidos en la poda durante el desarrollo son responsables del autismo ¹¹ y la esquizofrenia. ¹² Esta cualidad de «robusto pero frágil» es omnipresente, y el concepto resulta fundamental para entender cómo se organiza el cerebro.

La estrategia de diseño universal

Para Doyle, resulta evidente que la mayoría de los sistemas biológicos tienen una «arquitectura estratificada». Por lo tanto, cualquier intento de entender la experiencia consciente debe basarse en un sólido conocimiento del modo como el cerebro se organiza en capas. Puede que en un primer momento muchos investigadores, que han utilizado modelos cognitivos, no vean la diferencia entre «niveles» y «capas». Cuando solo existen niveles, todos los procesos son secuenciales (o, como dirían los ingenieros electrotécnicos, funcionan «en serie»), mientras que con una arquitectura estratificada el procesamiento se produce de forma simultánea (esto es, «en paralelo»). En el procesamiento mediante niveles, todos los pasos se realizan uno tras otro, como en una carrera de relevos: es necesario completar un nivel antes de que pueda iniciarse el siguiente. En el procesamiento en capas, en cambio, todos

los corredores pueden partir a la vez y dirigirse a lugares distintos. Este cambio de arquitectura implica grandes diferencias.

La arquitectura en capas es la estrategia de diseño clave para garantizar la robustez y la funcionalidad de los sistemas tecnológicos y biológicos organizados. Es simple, necesaria, potente y enormemente beneficiosa. Así, por ejemplo, un buen diseño en un sistema técnico, como un Boeing 777, o biológico, como nuestro cerebro, está configurado de tal manera que los usuarios ignoran en gran medida todas sus complejidades ocultas.¹³ Simplemente nos limitamos a subir al avión, inclinar el asiento hacia atrás y sacar un libro o pedir algo de beber. No pensamos en los 150.000 módulos de subsistemas del avión y en la función que realizan; ni tampoco lo hacen los pilotos. Ni siquiera sabemos que hay 150.000 módulos de subsistemas. Y si el lector se ha saltado el capítulo anterior, es posible que ni siquiera sepa qué es un módulo. De manera similar, casi ninguno de nosotros da demasiada importancia a su cerebro a menos que algo funcione mal. La complejidad de nuestro cerebro estratificado se halla tan bien escondida que después de dos mil quinientos años todavía estamos tratando de descifrarla. Tanto en el Boeing 777 como en nuestro cerebro, la arquitectura del sistema oculta su complejidad. Sea como fuere, ¿qué es una arquitectura estratificada?

El objetivo de un ingeniero es diseñar y construir cosas que funcionen de manera eficiente, efectiva y fiable, lo cual ya resulta bastante difícil cuando se construye una pérgola¹⁴ en el jardín; de modo que imagine si lo que hay que construir es la Ópera de Sídney. En un proyecto de ese tipo, no solo las diversas partes del edificio tienen que confluir para funcionar de manera eficiente, efectiva y fiable, sino también los ingenieros que deben trabajar juntos para construirlo. No hay una única persona que diseñe todos y cada uno de sus elementos; sin embargo, si se utiliza una estrategia de diseño errónea para organizar a los ingenieros, este enfoque basado en la colaboración de muchas manos puede llevar directamente al desastre.

De hecho, los propios ingenieros que diseñan sistemas complejos son en sí mismos sistemas complejos y están organizados de manera similar. Consideremos las diferentes estrategias para diseñar un Boeing 777 y su funcionamiento. Una posible estrategia es que, para diseñar el trozo del avión que se le ha asignado, cada ingeniero debe entender el trabajo de todos los demás. Luego, una vez diseñados los diferentes trozos, cada uno de los componentes tiene que depender de todos los demás para funcionar de manera adecuada; es decir, que todo debe estar integrado en serie. Eso implicaría que el ingeniero que diseña los asientos necesitaría saber todo lo que hay que saber sobre los motores, la sustentación y resistencia, el cristal de las ventanas, los sistemas de presurización, etc., y debería integrar plenamente la función del asiento con todos ellos. La construcción del avión no solo llevaría más tiempo, resultaría más costosa y requeriría un mayor número de expertos multidisciplinarios, sino que además sería vulnerable a más errores. Un asiento que no se inclinara hacia atrás no solo supondría una molestia: podría hacer caer el avión en picado.

La mejor estrategia es diseñar de forma independiente componentes que funcionen también de forma independiente (capas o módulos). Los diseñadores trabajan solo con la información que «necesitan saber»; el resto de la información se les oculta. En el mundo de la ingeniería, esto se conoce como «abstracción», es decir, la eliminación de detalles innecesarios (más abstracción = menos detalles), mientras que el concepto de «capas de abstracción» hace referencia a qué información está disponible y cuál se oculta. Las capas de abstracción no siempre requieren una jerarquía ni siquiera componentes fundamentalmente distintos. Un atlas del mundo tiene múltiples capas de abstracción, pero cada capa se da en la misma forma. La primera página exhibe un mapa del mundo. En él podremos ver rotulados los océanos, los continentes y quizá los principales ríos y cordilleras. Sin embargo, se ha omitido la mayor parte de la información: no hay países, ciudades, carreteras, arroyos ni colinas. Si pasamos la página nos

encontraremos con la siguiente capa de abstracción: un continente con varios países, capitales, ríos y montañas. Volvamos a pasar página y veremos un mapa de un solo país con mayor detalle, en el que aparecerán destacadas las principales carreteras y las ciudades más pequeñas. En cada nivel de abstracción podemos ver más y más detalles; se oculta menos información, por tanto. Pero más información no tiene por qué ser intrínsecamente mejor: si solo nos interesa el tamaño relativo de los diferentes océanos, no necesitamos saber que hay un sendero entre Roussillon y Fontaine-de-Vaucluse.

En un sistema complejo, en cambio, no se oculta simplemente la información. En este caso cada capa tiene una forma distinta, de modo que, para transferir la información necesaria entre diferentes capas, es necesario virtualizarla, es decir, abstraerla a una capa específica. Así, en el Boeing 777, al ingeniero que diseña los asientos solo se le proporcionará la información necesaria para hacer asientos: un conjunto de medidas estándar que permiten flexibilidad en el diseño del asiento pero a la vez constriñen el proceso de manera que todos los asientos encajen en todos los 777. No se proporciona información alguna sobre aerodinámica, ni sobre combustible; ni siquiera el número de asientos que tendrá el avión. He observado también que, al parecer, a los diseñadores de los asientos de los aviones tampoco les ha informado nadie de que hay personas que miden más de metro ochenta de estatura.

Aquí los asientos se convierten en módulos intercambiables. El funcionamiento del asiento no afecta a la capacidad de volar del avión. De modo que el ingeniero que diseña los asientos sabe más de aviones que nosotros, pero menos que los ingenieros que diseñan el fuselaje. Al mismo tiempo, los ingenieros responsables de las turbinas del avión no necesitan saber nada sobre el diseño de los asientos, pero, en cambio, disponen de mucha otra información.

La Madre Naturaleza descubrió esto hace tiempo y utiliza la misma estrategia en los organismos fruto de la evolución. Los diversos sistemas que integran nuestro cerebro han evolucionado para funcionar de manera

independiente. Así, por ejemplo, nuestro sistema auditivo funciona independientemente del olfativo. No obtiene información sobre olores, ni tampoco la necesita para procesar la información relativa al sonido. Podemos perder el sentido del olfato y seguir oyendo el zumbido de las abejas.

En una arquitectura estratificada, cada capa de un sistema opera de forma independiente porque cada capa tiene sus propios protocolos concretos; esto es: el conjunto de reglas o especificaciones que estipulan las interfaces o interacciones permitidas tanto en el seno de cada capa como entre capas distintas. Considere de nuevo el caso del ingeniero encargado de diseñar los asientos. Este puede soltarse cuanto quiera en sus diseños de asientos con tal de que se mantenga dentro del conjunto de medidas estándar, es decir, dentro del protocolo de la capa correspondiente a los asientos. El protocolo de una capa plantea una serie de restricciones, pero permite la flexibilidad dentro de estas.

Cada una de las capas que forman parte de una «pila» procesa la información que recibe de la capa inferior de acuerdo con su protocolo específico, y transmite el resultado a la capa superior y/o de vuelta a la inferior. La capa superior hace lo mismo según su propio protocolo específico, que puede ser similar al anterior o completamente distinto, y transmite los resultados de su procesamiento a las capas superiores. Ninguna capa «sabe» qué información ha recibido la capa anterior o qué procesamiento se ha producido. No es necesario, de modo que esa información se oculta (se abstrae). Los protocolos permiten que cada capa interprete solo la información que recibe de sus capas vecinas. La información generada por el procesamiento que tiene lugar dentro de la capa puede enviarse hacia arriba o hacia abajo. Solo hay una pega: una vez se crea una arquitectura estratificada, la información no puede saltarse ninguna capa. Así, una sexta capa no podría interpretar la información transmitida por la cuarta porque no cuenta con el protocolo adecuado para descifrarla, y, por lo tanto, necesita la intermediación

de la quinta. El objetivo de cada capa es servir a la capa inmediatamente superior a la vez que oculta los procesos de la inferior.¹⁵

He aquí un sencillo ejemplo de capas con protocolos: imagine que está en una fiesta con muchos invitados de diferentes países. Quiere hablar con la mujer china que aparentemente conoce a su hermana. Usted solo habla español, pero su pareja habla español y francés. La mujer china solo habla chino mandarín; sin embargo, su esposo habla chino y francés. Cada persona se convierte aquí en una de las capas que integran la pila de traducción, y cada capa utiliza su propio protocolo para transformar la información de entrada en una nueva versión de salida. Usted tiene el protocolo del español y transmite la información en español a su pareja, quien a su vez utiliza los protocolos del español y el francés para transmitírsela en este segundo idioma a la siguiente capa, el marido de la mujer china. Este comparte el protocolo del francés, pero también tiene el del mandarín, con el que transmite la información a su esposa. Luego esta última, a su vez, puede transmitirle información a usted a través de las diferentes capas; pero ni usted ni su invitada china pueden saltarse las capas francesas intermedias. La información puede fluir a través de las diferentes capas tanto en sentido ascendente como descendente, pero siempre debe ser procesada por la capa que dispone del protocolo adecuado para transmitirla a la siguiente capa. Si usted pretende crear su propia capa de transmisión del español al mandarín, no lo logrará en el transcurso de una cena.

Ahora bien, no hay duda de que puede crear un protocolo del español al mandarín si lo desea. Mucha gente lo hace. Se llama aprender un idioma, y no solo nosotros podemos hacerlo, sino también las máquinas. De hecho, los informáticos llevan años utilizando arquitecturas estratificadas, especialmente en el campo de la inteligencia artificial. Rodney Brooks, un ingenioso profesor de ciencias de la computación del MIT, propuso el concepto de «arquitectura de subsunción», que dominaría el campo de la robótica durante muchos años.

No es que los diccionarios aporten mucho sobre este concepto,¹⁶ pero la

idea es bastante simple. Un sistema, como una persona, un ordenador, un robot o una biblioteca, posee cierto conocimiento almacenado. Aparece una nueva información que viene a añadirse al conocimiento general del sistema en cuestión. En condiciones ideales, dicha información es «subsumida» o absorbida por la ya existente sin alterar las cosas. Esta arquitectura es exactamente la que necesita un robot.

Brooks sabía muy bien que los robots no superaban la prueba de subsumir la nueva información porque hace veinte años, cuando propuso la idea, estos se quedaban atascados cada vez que se topaban con prácticamente cualquier objeto que no se hubiera especificado previamente en su programación que debían evitar (pongamos por caso, un postre casero). Eran incapaces de adaptarse a un entorno cambiante. Sin embargo, si un robot estuviera dotado de una arquitectura de subsunción, ello le permitiría realizar cambios progresivos añadiendo una a una nuevas capas a las ya existentes. Cada una de dichas capas podría ejercer su influencia en las capas inferiores y, de ese modo, quedaría subsumida en el conjunto de la arquitectura general. Como resumía Harold Pashler en su *Encyclopedia of the Mind*: «Una idea clave es que el sistema en su conjunto no construye representaciones integradas del mundo; antes bien, las señales sensoriales se procesan de forma distinta en cada nivel a fin de implementar correspondencias relativamente directas y específicas para cada comportamiento entre los datos de percepción y las señales motoras requeridas para controlar los actuadores del robot». ¹⁷ Eso significa que hay sistemas muy específicos repartidos por toda la arquitectura que gestionan de manera precisa los retos cotidianos del robot; sistemas que son a la vez rápidos, eficientes y útiles. No hay un sistema central que cambie la pauta de respuesta del robot para afrontar cada nuevo reto; en su lugar, se agrega una instrucción específica que aborde dicho reto; y, con el tiempo, los ingenieros van añadiendo cada vez más instrucciones para abordar cada vez más perturbaciones del entorno. No hay una gran unidad central tratando de solucionarlo todo, sino nuevas capas que se van añadiendo a medida que se

presentan los problemas, lo cual suena muy parecido a todos aquellos módulos que se combinaban en la arquitectura estratificada; de hecho, un módulo puede ser una capa en sí mismo o bien una capa puede estar integrada por un grupo de módulos. Cuando hablamos de módulos integradores al final del capítulo 4, estos eran módulos de procesamiento que formaban una capa en lo alto de una pila. Dicha capa recibía información de una capa anterior y la procesaba, de acuerdo con su protocolo, para producir algo más complejo, quizá incluso una teoría de la mente o la conciencia de sí.

La estratificación permite flexibilidad. Implementar actualizaciones en un sistema basado en capas es sencillo porque únicamente tienen que producirse cambios en una determinada capa concreta, sin necesidad de alterar las demás capas. Y cuando algo sale mal, la fuente del error resulta identificable. No es necesario reparar ni eliminar todo el sistema, sino solo la capa o las partes de esta que no funcionan. Como nos vestimos con distintas capas de ropa, si se nos rasga la camisa basta con cambiarla; no tenemos que cambiarnos también de pantalones. En el caso de nuestro cerebro puede que no resulte tan fácil cambiar las piezas, pero si algo va mal no tenemos por qué perder todo el sistema.

La neurociencia de la estratificación

La belleza de la arquitectura estratificada reside en el hecho de que, al ocultar información, resuelve un problema para el usuario de sistemas complejos. En nuestro iPhone, por ejemplo, la capa superior se conoce como «capa de aplicación»; se hace así para que no haya necesidad de saber o de entender cómo funcionan las demás capas del sistema. De ese modo, no tenemos que calcular los protocolos de asignación de memoria de nuestro iPhone cada vez que queremos enviar un mensaje de grupo o hacer una foto. De manera similar, debemos estar agradecidos por no tener que entender cómo funciona el cerebro para poder usarlo. No sabemos cómo nuestro almuerzo se convierte en

energía; simplemente lo tomamos y seguimos adelante. Y lo mismo ocurre con nuestra vida mental: no tenemos ni la menor idea de cómo hacemos cualquier cosa. Por ejemplo, ¿sabemos cómo conseguimos tocarnos la nariz? Saberlo en términos de generar los mensajes neuronales que enviamos a los músculos está completamente fuera de nuestro ámbito de conciencia y de conocimiento. Al igual que el piloto de un 777 maneja un programa informático para hacer volar el avión, nosotros manejamos nuestra mente, la «capa de aplicación» del cerebro, para generar acciones, es decir, nuestro comportamiento. Pero ¿de verdad posee el cerebro una arquitectura estratificada? ¿O esta es un mero formalismo que en realidad no tiene cabida en los sistemas biológicos reales?

Como suele ocurrir, el descubrimiento de una nueva perspectiva o idea generalmente se traduce en el descubrimiento de que otros científicos ya habían tenido ideas similares, a veces años atrás. ¿Cuántas veces necesitaremos aprender que las ideas humanas son exactamente eso: ideas humanas; es decir, ideas que muchos humanos han tenido a lo largo de la historia? Analicemos el caso de Tony Prescott, Peter Redgrave y Kevin Gurney, de la Universidad de Sheffield. Los tres están muy versados en neurociencia, robótica y ciencias de la computación, que utilizan con una ilimitada inteligencia. En un importante artículo sobre estratificación que escribieron hace casi veinte años ¹⁸ nos encaminaron hacia donde hoy nos encontramos. La historia empieza con el gran neurólogo británico del siglo XIX John Hughlings Jackson. Este era un médico brillante en todos los aspectos, lo que podríamos llamar un jugador de primera división. Por desgracia, jugaba en tercera cada vez que cogía la pluma, ya que sus escritos resultaban prácticamente impenetrables, aunque por fortuna ha habido otros jugadores de primera que han hecho que su trabajo resulte comprensible al mundo.

Darwin había inspirado al mundo científico y médico, y Jackson estaba plenamente comprometido con sus ideas. El cerebro, fruto de la selección natural, era una máquina sensoriomotora, y cada especie había desarrollado su propio conjunto de habilidades. En los humanos, las capas superiores eran las

más sofisticadas a la hora de coordinar una acción, pero tanto estas como las inferiores tenían incorporadas las habilidades básicas. Por ejemplo, un gato o una rata al que se le extirpe la corteza cerebral seguirá exhibiendo diversos comportamientos motivados, como andar, acicalarse, comer y beber. Pero sin las capas superiores desaparecen ciertos comportamientos más complejos. En palabras de Prescott y sus colegas de Sheffield:

[Jackson] dividió el sistema nervioso en centros inferiores, medios y superiores, y postuló que esa secuencia representaba una progresión de lo «más organizado» (más fijo) a lo «menos organizado» (más modificable), de lo «más automático» a lo «menos automático», y de lo más «perfectamente reflejo» a lo menos «perfectamente reflejo». Esta progresión contempla un incremento en la competencia de una manera que hoy podríamos entender como una descomposición conductual: los centros superiores se ocupan del mismo tipo de coordinaciones sensoriomotoras que los inferiores, aunque de una forma más indirecta. ¹⁹

Jackson supo ver de inmediato las implicaciones de su concepción estratificada, y sugirió que debía de haber «disociaciones» —un término que él introdujo en la neurología— que hacían que determinadas lesiones cerebrales específicas produjeran determinados tipos concretos de déficits conductuales. Elimínense las capas superiores, y solo las inferiores podrán responder. Y responderán solo con su limitado nivel de capacidad, tal como se describía en el caso del gato decorticado.

Sistemas estratificados susceptibles de evolucionar

Obviamente, todo este trabajo pionero obligaba a plantearse la cuestión de si el cerebro se desarrolla o no de manera estratificada en una escala de tiempo evolutiva. ¿Se van añadiendo las distintas áreas cerebrales de forma lenta pero segura, y existen evidencias de ello derivadas de los estudios de anatomía comparada? De hecho sí, y ahí es donde destaca Prescott, que explica una larga, compleja y fascinante historia: el proceso evolutivo del sistema nervioso actual de todos los vertebrados se inició hace más de

cuatrocientos millones de años con el plan básico de la médula espinal, el rombencéfalo, el mesencéfalo y el prosencéfalo. Con el transcurso de los milenios, el prosencéfalo fue incorporando módulos y capas que aportaban nuevas funcionalidades, en lugar de limitarse simplemente a desarrollar las antiguas. Por ejemplo, en la medida en que las extremidades iban haciéndose más capaces de manipular cosas se requería una mayor extensión neuronal para que las capas modulares ejercieran el control de los nuevos manipuladores periféricos, más conocidos como dedos. Esas nuevas vías neuronales están claramente presentes en los vertebrados dotados de dedos, pero no en los que carecen de ellos. Asimismo, y tal como predijo Jackson, las lesiones del prosencéfalo alteran el funcionamiento de una serie de módulos que gobiernan la motricidad fina de la mano, pero no perturban el de otros responsables del control motor, más básico, del brazo.

La capacidad de evolucionar es algo bueno para cualquier población animal, ya que permite adaptarse a los nuevos desafíos, y se define como la capacidad de un organismo de generar variaciones fenotípicas (rasgos observables) que pueden heredarse.²⁰ Si, como resultado de la selección natural, se selecciona un rasgo determinado, este se transmitirá a la próxima generación. Un ejemplo bien conocido, observado en las islas Galápagos, es la diversidad de tamaños de picos —desde los más pequeños hasta los más grandes— que se dan en el pinzón terrestre.²¹ Sin embargo, la teoría darwiniana de la selección de las variaciones hereditarias planteaba (entre otros) este enigma: ¿de dónde vienen exactamente todas esas variaciones y cómo se generan? La razón que habitualmente se aduce —que se deben principalmente a mutaciones genéticas aleatorias— puede ser parte de la explicación, pero no toda. Este enigma ha desconcertado a los biólogos durante años.

Entonces aparecieron el biólogo de Harvard Marc Kirschner y su colega de Berkeley John Gerhart.²² Ambos se preguntaron si las criaturas modernas tienen mecanismos celulares y de desarrollo con la característica de lo que se

conoce como «evolucionabilidad» o capacidad evolutiva. En otras palabras: ¿tienen la capacidad de generar variaciones fenotípicas hereditarias?, ¿y esa característica en sí misma es objeto también de la presión selectiva? Es decir, ¿los sistemas biológicos que producen más variaciones fenotípicas susceptibles de transmitirse a sus descendientes son los que tienen más probabilidades de ganar la carrera armamentística de la evolución?

En todo el reino animal existe una gran diversidad en cuanto a forma del cuerpo, organización de los tejidos, desarrollo y fisiología. Al mismo tiempo, muchos procesos esenciales, como las vías de transmisión de señales bioquímicas y celulares, así como los circuitos que regulan la expresión génica, son iguales en todos ellos. Los animales compartimos algunos procesos esenciales con las plantas, los hongos o el moho mucilaginoso; por ejemplo, empleamos las mismas enzimas para regular la división celular. También compartimos otros procesos esenciales —el metabolismo y la replicación— con formas de vida tan inferiores como las bacterias. ¿Por qué? Pues simplemente porque compartimos muchas de las mismas secuencias genómicas. Aunque algunos biólogos creen que esos procesos esenciales limitan la evolución, Kirschner y Gerhart no son de la misma opinión. De hecho, ellos piensan justo lo contrario: creen que la razón de que compartamos tantos procesos esenciales, de que estos hayan estado ahí durante los últimos quinientos treinta millones de años y de que se hayan revelado tan fructíferos, es que no solo confieren flexibilidad (en lugar de limitarla) sino que además han permitido que las variaciones de mayor éxito se transmitan a la descendencia. La flexibilidad conferida por estos procesos esenciales se traduce concretamente en la variación fenotípica de aquellos procesos que resultan ser frágiles frente a los cambios ambientales. Por lo tanto, los procesos esenciales constituyen una restricción que de hecho ha proporcionado flexibilidad evolutiva frente a la incertidumbre ambiental.

Si al lector le parece que todo eso de las «restricciones que flexibilizan» suena a protocolos en una arquitectura estratificada, acierta. Doyle y Alderson

lamentan que la mayoría de los biólogos no hayan sabido apreciar plenamente el papel que desempeña la arquitectura estratificada en la producción de variaciones. Quizá la arquitectura estratificada, que es un rasgo omnipresente en los sistemas biológicos, evolucionó porque su capacidad de generar variaciones dentro de un conjunto de restricciones era robusta frente a la competencia y fue seleccionada por esta: ¡estratificarse o morir!

Vínculos que liberan

Identificar una arquitectura estratificada es una cosa, pero otra muy distinta es averiguar cómo sus capas se comunican entre sí. Cada una de ellas debe procesar la amplia variedad de información que recibe y darle una nueva forma que sea interpretable por la siguiente capa. Las principales restricciones de una estructura estratificada se dan en los procesos que conectan las diferentes capas entre sí.²³ Una forma ingeniosa de visualizar esta característica de la arquitectura es imaginarla como una pajarita o un reloj de arena, donde el protocolo está representado por el nudo o el estrechamiento central desde el que se despliega la información de entrada y de salida. En nuestro anterior ejemplo del ingeniero encargado de diseñar los asientos del avión, las especificaciones técnicas que integran el protocolo de la «capa de asientos» constituyen el nudo de la pajarita. La información entrante en el protocolo la integran todo tipo de posibles materiales de construcción, formas, colores, etc., mientras que el producto saliente pueden ser toda una serie de asientos de diversos materiales, diseños y colores; pero todos ellos tienen que ajustarse a las especificaciones y a la función del protocolo. En conjunto, el sistema está constreñido, pero a la vez flexibilizado, o «deconstreñido» (por utilizar una palabra *ad hoc*).²⁴ Un determinado número de entradas pueden dar lugar a diferentes tipos de salidas, puesto que existen múltiples formas de realizar la tarea de una capa. Si uno lo piensa bien, parece cosa de magia. Cuando se observa el comportamiento de un robot dotado de la adecuada

arquitectura estratificada, casi parece que este sistema —estrictamente definido y con solo respuestas preestablecidas— «piense» realmente, en el sentido emocional del término. La arquitectura permite al sistema ser más flexible.²⁵

El hecho de que el protocolo de una capa constriña y a la vez flexibilice un sistema resulta crucial. Me gustaría insistir en ello, así que ahí va otro ejemplo. Piense de nuevo en sus capas de ropa y en toda la gama de posibilidades que cada capa permite. Por ejemplo, la capa de más abrigo (con su protocolo constrictivo: debe atrapar el calor corporal) tiene toda clase de prendas como material de entrada, y muchos posibles atuendos como producto de salida. Estos podrían ser, por ejemplo, un capote de piel de oso y unos pantalones de lana, un abrigo de lana merina con un suéter de cachemira y unos pantalones de piel de oveja, o una chaqueta de forro polar con un chaleco de visón y unos pantalones de goma isotérmicos. Puede abrocharse con cremallera o meterse por la cabeza; puede ser una pieza única, tipo mono, o bien estar integrado por dos piezas distintas. Puede tener cuello de cisne y mangas y tobilleras elásticas, o no. Puede ser de cualquier color o tamaño. Aunque la capa de abrigo tiene un protocolo que la constriña (debe atrapar el calor corporal), ese mismo protocolo también la deconstriña en diversos aspectos, lo que permite una gran variedad de opciones. En términos darwinianos, se da una selección a partir de la variación. Es fácil comprender cómo esa flexibilidad ha permitido evolucionar a esta capa desde una sencilla piel de animal con la que cubrirse hasta una chaqueta de forro polar de talla L de color magenta con cierre de cremallera y capucha. Y bolsillos. Esto ilustra el que podría ser el rasgo más importante de la arquitectura estratificada: que posibilita el cambio a lo largo de una prolongada escala de tiempo, pasando de la indumentaria de los Picapiedra a un traje de Armani o un vestido de Valentino.

Pese a la flexibilidad de la arquitectura estratificada, esta también tiene sus inconvenientes. Tomemos como ejemplo las permutaciones mencionadas en la

ropa de abrigo y preguntémosnos ahora cuáles de ellas están «de moda». Este es un problema más difícil. Cuanto más específico sea el protocolo, mayor será la restricción. En ese aspecto, un sistema unificado que no opere en capas tiende a funcionar de forma más eficiente, puesto que no tiene protocolos independientes para cada una de sus funciones. Recuérdese que un protocolo es un conjunto de reglas o especificaciones que estipulan las interfaces o interacciones permitidas tanto dentro de cada capa como entre capas distintas. Sería más fácil limitarse a diseñar un mono de un material caro que te mantenga caliente y seco y a la vez resulte confortable, una prenda única que sea flexible y liviana. Cabría sin problemas en nuestro equipaje de mano y resultaría fácil de poner y quitar. Hasta podría darnos un aspecto estilizado, y encima estar a la moda. Así que solo necesitaríamos esa única prenda para el resto de nuestra vida. ¡Genial!

Sin embargo, una estructura de funcionamiento unificada no es la solución ideal para un sistema complejo, puesto que bastaría una pequeña disfunción para dar al traste con todo, y además no resulta fácil aplicar actualizaciones. Si nos enganchamos la pernera de nuestro mono ideal con un clavo, acabará deshilachándose toda la prenda. Al menos en el caso de un sistema estratificado dañado contamos con algunas medidas de respaldo, y resulta más fácil y más barato de reparar o de reemplazar. Además, ¿qué ocurre si aparece un material que resulta más transpirable? Para poder aprovecharlo tendremos que desechar toda la prenda y hacernos con una nueva, lo cual, obviamente, es costoso. Se necesita más tiempo, energía y recursos para mantener un sistema unificado; es decir que, aunque pueda ser más eficiente, la contrapartida es que también resulta más costoso y no tan robusto. Dado que cada capa puede proveer una amplia gama de funciones diversas, el sistema en su conjunto tiene una mayor flexibilidad, lo que le proporciona una gran ventaja a la hora de afrontar un entorno cambiante. Este tipo de diseño resulta ideal en un sentido evolutivo porque el número de vulnerabilidades del sistema es limitado, mientras que abundan las oportunidades de diversificación. A medida que el

entorno va cambiando con el tiempo, tales sistemas pueden adaptarse con mayor facilidad. En términos generales, una arquitectura estratificada es la solución ideal para sistemas complejos, puesto que es fácil de reparar, resulta menos costosa y más flexible, y posee una mayor capacidad evolutiva.

Ahora bien, los sistemas complejos estratificados no son inmunes a las disfunciones de protocolo. Cuando el sistema se rompe, es objeto de bloqueo o incluso es interceptado, pueden producirse fallos, algunos de ellos de consecuencias catastróficas. Por ejemplo, si las costuras de nuestros pantalones de lana se descosen debido a un fallo de protocolo en la capa de costura, los pantalones dejarán de ser funcionales: daría igual que lleváramos una falda de paja. Si en el sistema biológico que es nuestro cuerpo los protocolos del sistema inmunitario son interceptados, podemos terminar padeciendo una enfermedad autoinmune. Dado que los sistemas complejos tienen numerosos componentes y capas de subsistemas, pueden darse interacciones que son muy difíciles de predecir. Por ejemplo, un pequeño fallo en un componente que no afecte en gran medida al rendimiento local de dicho componente puede verse amplificado cuando interactúa con otros componentes y afecta el rendimiento general del sistema. Esta interacción imprevista puede provocar un fallo del sistema.²⁶

Los daños producidos en los protocolos pueden comprometer incluso al sistema más robusto, pero la ventaja es que, en general, hay pocos puntos vulnerables importantes. Y lo que es aún mejor: el sistema puede seguir funcionando, aunque sea renqueando, a pesar de su déficit, mientras que en un sistema unificado un ataque a cualquier componente puede comprometer todo el sistema.

¿Por qué no podemos limitarnos simplemente a eliminar esos errores del sistema? ¿Por qué no podemos deshacernos de esos puntos vulnerables? El problema es que cada nueva estrategia que resuelve un problema introduce invariablemente alguna nueva vulnerabilidad, que a su vez tiene que repararse. El aumento de la robustez es una carrera armamentística que da como

resultado que los sistemas vayan agregando cada vez más capas, incrementando con ello su complejidad. La evolución de nuestro complejo cuerpo y nuestro complejo cerebro a partir del primitivo caldo químico, y la del 777 a partir de una serie de tuercas y tornillos, y de la bicicleta de Orville Wright, han sido simplemente el resultado de una carrera en la que se ha ido añadiendo una capa tras otra de características robustas destinadas a combatir determinadas fragilidades. Es como la Reina Roja en *A través del espejo*, que corre cada vez más deprisa para permanecer en el mismo sitio. Entonces, ¿estamos condenados? Una posible contraestrategia consiste en introducir redundancias en el sistema.

Todos los cerebros llevan a Roma

Por más que el protocolo de una capa constriña el producto de esta última limitando sus posibilidades, no dicta cuáles de esas posibilidades se materializarán en dicho producto. Hablar de constricciones que deconstruyen no es lo mismo que hablar de causalidad. Una constricción puede limitar el número de resultados, pero no causa el resultado. Si nos vestimos para una fiesta, nuestra indumentaria se verá constreñida por la ropa que tengamos en el armario (y posiblemente por nuestro concepto de lo que es un atuendo de fiesta socialmente aceptable), pero esa constricción no nos obliga a llevar una determinada prenda concreta. Sigue habiendo mucho donde elegir. Las constricciones de protocolo que deconstruyen no dictan el resultado. Partir del supuesto de que sí lo hacen puede causar un problema para quienes no son conscientes de que están tratando con protocolos en el marco de una arquitectura estratificada, que pueden verse inducidos a creer que, observando un determinado comportamiento, pueden predecir la pauta de activación neuronal, es decir, el «estado cerebral» que lo ha producido. El error de esta forma de pensar ha quedado bien patente gracias al trabajo de la brillante

neurocientífica Eve Marder,²⁷ que ha estado estudiando las tripas de la langosta.

Marder se ha dedicado a estudiar la «capa digestiva» de la langosta, examinando las contracciones del intestino. Para ello, aisló y estudió todas y cada una de las neuronas y sinapsis —incluyendo los efectos de sus neurotransmisores— involucradas en la motilidad intestinal de la langosta. Así como nosotros podemos tener un millón de posibles atuendos en nuestro armario (combinando la ropa en todo tipo de variaciones extravagantes, como llevar calcetines en las manos o ponerse una falda encima de unos vaqueros), Marder descubrió que en ese pequeño intestino hay dos millones de posibles combinaciones reticulares. Pero, como ocurre con nuestro atuendo, las constricciones de protocolo reducen los posibles resultados: solo un pequeño porcentaje de ellos funcionan. Así, por ejemplo, normalmente no llevamos los calzoncillos encima de los pantalones ni la chaqueta debajo del vestido de cóctel, aunque podríamos hacerlo. Al igual que con nuestra indumentaria, un pequeño porcentaje de un par de millones sigue siendo una gran variación. De hecho, realmente funcionaba entre un 1 y un 2 %; es decir: entre 100.000 y 200.000 ajustes de ese puñado de neuronas se traducirán en el mismo comportamiento exacto en cualquier momento dado. Había múltiples formas de realizar la tarea de la capa de motilidad, del mismo modo que hay diversas formas de realizar la tarea de agrupar prendas para formar un atuendo. Este es un ejemplo de lo que se conoce como «realización múltiple», la noción de que, en lo que se refiere a la activación de neuronas, hay muchos caminos (aunque no todos) que llevan a Roma. Es decir, que la misma propiedad, estado o acontecimiento mental puede implementarse mediante diferentes pautas de activación neuronal. Esto puede parecer un desperdicio de tiempo evolutivo o de energía bioquímica, pero lo que realmente significa es que si una vía neuronal se deteriora, otra puede reemplazarla.

Los sistemas estratificados minimizan los costes en recursos desarrollando componentes adaptables que pueden servir a múltiples propósitos. Así, por

ejemplo, en la capa bioquímica del cerebro hay muchas proteínas que cumplen múltiples funciones relacionadas con vías de transmisión de señales y bucles de realimentación a fin de regular el sistema en diversas capas.²⁸ Así el sistema ahorra energía, puesto que no necesita desarrollar varios componentes únicos para cada capa.

Otro elemento reductor de costes es el control de daños que proporciona el procesamiento en paralelo. Por ejemplo, en los organismos multicelulares existe una capa celular en la que las células se metabolizan individualmente y cada una de ellas bulle de actividad independientemente de las demás, y existe también una capa de tejido formada por esas mismas células que en este caso trabajan juntas realizando la tarea propia del tejido y siguiendo el protocolo propio de ese último. En la capa celular, cada célula lleva a cabo sus propios protocolos; estos pueden ser idénticos a los de otras células del mismo tejido, pero cada una de ellas bulle de actividad de manera independiente. Si una célula resulta dañada, solo se requerirán los recursos necesarios para reavivar esa única célula en lugar del tejido entero. Asimismo, si una célula queda destruida y es imposible repararla, lo más probable es que el tejido siga funcionando, ya que habitualmente la ausencia de una sola célula que funcione de manera independiente se puede pasar por alto. Sin embargo, como ocurre en todos los sistemas estratificados, también este puede verse comprometido si falla un componente que resulta esencial para la transmisión de información de una capa a otra. En el caso del tejido biológico, si los tipos de proteínas que interconectan las células son defectuosos, es posible que la información no pueda transferirse de la capa de las células individuales a la capa del tejido, y el sistema entero podría venirse abajo. Aunque no es perfecto, el diseño estratificado de los sistemas biológicos resulta ventajoso en cuanto que minimiza el número de puntos en dichos sistemas que, de verse atacados, provocarían un daño catastrófico, y limita también los efectos de los ataques en otros puntos.

Kirschner y Gerhart postulan que los rasgos que comportaban una mayor

flexibilidad sentaron las bases de los organismos con un desarrollo más complejo, puesto que los sistemas más dinámicos tienden a resultar menos vulnerables a las mutaciones letales. Las poblaciones de organismos pueden evolucionar gracias a la variación fenotípica que surge de una arquitectura estratificada. En consecuencia, es posible que la capacidad evolutiva de los sistemas estratificados haya contribuido en gran medida al éxito en materia de supervivencia a lo largo del tiempo.

De los cromosomas a la conciencia: una arquitectura estratificada

Aquí es donde entramos en terreno resbaladizo. Aunque entendemos que los sistemas biológicos complejos tienen una arquitectura estratificada, desconocemos aún el papel general de cada capa, así como la totalidad de sus funciones y su dinámica. Algunos sistemas, como las bacterias, resultan más fáciles de entender que otros, como —sí, lo ha adivinado— el cerebro. En un sistema biológico, lo que Doyle denomina «capas de composición» —las capas evolutivamente más antiguas— constituyen las piezas más básicas del sistema; sus elementos integrantes son las partículas subatómicas, los átomos y las moléculas, cada uno con sus propios protocolos. Otras capas incluirían, por ejemplo, una dotada de protocolos que describan las interacciones entre las distintas moléculas y dentro de cada una de ellas; otra con un protocolo capaz de generar interacciones dinámicas, es decir, aquellas que producen cambio o progresión; y una serie de capas de control con protocolos que utilicen la realimentación para ajustar la respuesta del sistema. Los protocolos de control guían el comportamiento del sistema frente a diversas perturbaciones internas y externas. Y lo que sí sabemos es que los sistemas de control del cerebro humano en fase de maduración tardan más que muchos otros sistemas en llegar a estar a pleno rendimiento. Si no está convencido de ello, intente perturbar a un adolescente.

¡Contrólese!

El control crea orden y precisión en un sistema, evitando que este actúe de forma aleatoria. Imagínese qué ocurriría si nuestras neuronas se activaran al tuntún: no podríamos meternos el tenedor en la boca, y mucho menos caminar por la cuerda floja. Los sistemas de control se pueden orientar hacia un control óptimo, que optimiza el rendimiento en un escenario medio, neutro en materia de riesgo, o hacia un control robusto, que tiene en cuenta el factor riesgo y optimiza el rendimiento con vistas a los escenarios más desfavorables. Los sistemas de control óptimos, como su propio nombre indica, constituyen la mejor solución frente a un determinado reto conocido; sin embargo, pueden resultar arbitrariamente frágiles frente a otras incertidumbres.²⁹ Por lo tanto, lo que actualmente predomina en los sistemas tecnológicos es el control robusto (que, no obstante, se mantiene oculto, y solo se revela ante lo que no sucede habitualmente: accidentes, averías, etc.).³⁰ El Boeing 777 es capaz de atravesar tormentas y salir airoso porque sus sistemas de control están diseñados para ser robustos frente al mal tiempo en lugar de estar optimizados para los cielos azules y despejados.

La mayoría de los neurocientíficos consideran que el cerebro está dotado de sistemas de control óptimos, una opinión de la que discrepan el neurocientífico, ingeniero y médico Daniel Wolpert y sus colegas, que creen, por el contrario, que la versión robusta explica mejor el control de la motricidad humana.³¹ Los sistemas de control robustos tienen unos límites estrictos en cuanto a robustez y eficiencia,³² y deben afrontar disyuntivas entre ellos: velocidad frente a precisión, velocidad frente a flexibilidad, flexibilidad frente a eficiencia, velocidad frente a coste... Y precisamente este tipo de disyuntivas están bien documentadas en todo tipo de procesamientos conscientes y no conscientes.³³

Para Wolpert, el control de la motricidad es lo único que importa. Como científico, él forma parte de una serie de autodenominados «chovinistas motores», una estirpe que incluye a figuras como el premio Nobel sir Charles

Sherrington, que escribió: «El propósito de la vida es un acto, no un pensamiento», o Roger Sperry, que nos alentó a «ver el cerebro objetivamente como lo que es, a saber, un mecanismo que gobierna la actividad motora». ³⁴ Al fin y al cabo, es la acción, no la reflexión, la que pone la comida en la mesa y cuece un bizcocho en el horno. La acción permitió a nuestros antepasados sobrevivir y reproducirse. Wolpert, probablemente el líder actual de este grupo, sostiene que la única razón por la que tenemos un cerebro es para poder movernos de manera adaptable. ³⁵ Puede que esta afirmación a usted le resulte irritante, pero considere el hecho de que el corazón es un músculo sin cuyo movimiento no podemos vivir. Asimismo, es la motricidad la que obtiene el alimento, lo mastica y lo digiere. Sin alimento, el cerebro no puede funcionar, y, desde luego, no puede producir los aspectos creativos de la vida —la literatura, el arte y la música—, que, en cualquier caso, se quedarían atrapados en el cerebro de no existir la motricidad para transmitirlos al mundo exterior a través del habla, la escritura, los gestos manuales o las expresiones faciales. Tenemos que contemplar esta idea con la perspectiva que nos aporta. Si nuestro cerebro ha evolucionado como sistema de control motor del cuerpo, entonces pensar, planificar, recordar, emplear los sentidos, etc., son simples herramientas, complejidades añadidas en una arquitectura estratificada que ha evolucionado para incrementar la robustez del control motor en entornos cambiantes e inciertos. Y eso vale también para el aprendizaje y la cognición. Y, como ocurre en todos los casos, estas capas evolutivamente más recientes traen consigo sus propias fragilidades.

Si tocamos involuntariamente un quemador caliente, nuestra respuesta es un reflejo automático: retiramos el dedo con un movimiento rápido antes de llegar a sentir el dolor de manera consciente; es lo que se conoce como «control por realimentación», que en este caso se produce en el nivel del sistema nervioso periférico. Las neuronas de la médula espinal —rápidas, gruesas y aisladas (y, por eso mismo, costosas)— provocan el alejamiento inmediato del estímulo doloroso sin que intervenga para nada el cerebro

consciente. El reflejo es automático, rápido, costoso en términos energéticos y oculto a la percepción consciente; pero no es flexible. Apartamos la mano de un tirón, en un único movimiento brusco: no es que esta se ponga precisamente a aletear como una mariposa. Tras una breve pausa, sin embargo, entran en juego los nervios especializados —lentos y finos—, que son los que nos proporcionan la información consciente sobre el origen del movimiento: «¡Ay, me duele el dedo!». ¿Y qué ocurre después? La lenta cognición consciente produce toda una serie de respuestas diversas con respecto a qué hacer para aliviar el dolor, tanto en ese momento como en el futuro: podemos metérselo en la boca, introducirlo en un recipiente con agua helada o frotarlo con aloe vera; y nos proponemos no volver a tocar un quemador caliente. La cognición es precisa, flexible y energéticamente barata, pero lenta. A veces podemos disponer de tiempo, pero en situaciones inciertas la lentitud puede resultar letal.

Podemos concebir el aprendizaje y la cognición como capas de control sofisticadas que han evolucionado con el objetivo de ser robustas frente a perturbaciones futuras planificando la respuesta a un estímulo que aún no se ha producido. En parte logran ese objetivo utilizando la realimentación de experiencias anteriores con un estímulo similar (memoria), pero dicha realimentación puede traducirse en algo más que el mero hecho de adaptarse al estímulo: con el tiempo también puede producir cambios en el protocolo de una capa; es lo que denominamos «aprendizaje».

Los humanos, los animales y algunos otros organismos pueden prepararse para los acontecimientos futuros mediante diversos mecanismos de aprendizaje. Si un animal se tropieza con un alimento que tiene mal sabor, aprenderá a evitar ese estímulo en el futuro. Sabemos que ha habido aprendizaje cuando un mismo estímulo («¡Caliente! ¡No tocar!») suscita una respuesta distinta en un organismo a lo largo del tiempo. El protocolo cambia, por ejemplo, de «Comer pájaros» a «Comer pájaros excepto cuervos», o de

«Investigar las cosas por medio del tacto» a «Investigar las cosas por medio del tacto, excepto las cocinas».

Si bien nacemos con algunos comportamientos automáticos, como el reflejo instintivo del dolor, otros se aprenden. El aprendizaje puede hacer que un determinado comportamiento «descienda un peldaño», pasando de la capa de control consciente, sofisticada pero lenta, a la capa inconsciente, rápida y automática. Así, por ejemplo, cuando practicamos nuestro swing de golf, hacemos una predicción de dónde irá a parar la bola a partir de nuestros recuerdos de swings anteriores. Una vez lanzada, también incorporamos la realimentación que nuestra visión nos proporciona acerca de dónde ha ido a parar realmente: ¡corto!; de modo que reajustamos un poco el swing y volvemos a intentarlo con algo más de brío. Ahora ha estado bien de largo, pero se ha desviado un poco hacia un lado; así que seguimos practicando. Con la suficiente práctica, la bola terminará aterrizando casi todas las veces justo donde queremos (siempre que no haya perturbaciones externas, como una ráfaga de viento o que un amigo nos cuente un chiste justo cuando estamos a punto de lanzar la bola; o perturbaciones internas, como que nos entre sed o nos dé un calambre muscular, o que nos venga a la cabeza algún pensamiento intempestivo sobre cualquier cosa). Llegados a ese punto ya no tendremos que pensar en controlar todos nuestros movimientos: estos se habrán hecho automáticos.

Ser robusto frente a futuras perturbaciones también implica planificar cosas que puede que no se hayan experimentado antes. Planificamos utilizando modelos internos para simular el futuro; recordamos acontecimientos pasados y re combinamos los recuerdos de diversas formas a fin de producir una serie de planes para adaptarnos a posibles circunstancias futuras. Por lo tanto, el protocolo para simular planes de nuestra capa de control consciente y sofisticada es una constricción que deconstruye. Es lo que hacemos, por ejemplo, cuando prevemos qué ropa vamos a llevarnos en nuestro viaje al norte. Para empezar, predecimos a qué podríamos tener que enfrentarnos, y

simulamos cómo nos sentiríamos partiendo de nuestros recuerdos de lo que hemos experimentado en el pasado en situaciones similares. Recordamos que entonces hubo un desajuste entre lo que queríamos sentir y lo que realmente sentimos, de modo que, para adaptarnos a esa circunstancia, hacemos que nuestra indumentaria resulte más robusta frente al clima frío y húmedo. A medida que las experiencias y los recuerdos que estas producen se van acumulando, aumenta la información de la que dispone el protocolo de planificación, que podrá combinar en una gama más amplia de escenarios futuros. Cuantas más experiencias hayamos tenido, más opciones podrá simular nuestro cerebro.

Llevar el concepto de estratificación a un ente biológico tan tremendamente complejo como usted y yo equivale de hecho a plantear toda una perspectiva acerca de cómo concebir el posible funcionamiento de ese sensible ente biológico. Descomponer las cosas en capas que interactúan entre sí proporciona al ingeniero un marco para pensar en cómo construir un cerebro. Aunque todavía nadie está cerca siquiera de hacer tal cosa, esa perspectiva sí sirve de guía a los neurobiólogos, que trabajan arduamente para estudiar neuronas individuales, o pequeños circuitos de neuronas, a la hora de reflexionar sobre sus hallazgos. Y abre un camino para explicar cómo un sistema complejo integrado por numerosas piezas locales se puede organizar para realizar una tarea de tanta envergadura como diseñar un teatro de la ópera.

6

El abuelo está demente, pero conserva la conciencia

No importa cuán hermosa sea tu conjetura. No importa lo inteligente que seas... Si no concuerda con el experimento, está equivocada.

RICHARD FEYNMAN

La conciencia es resiliente y difícil de eliminar. Lo aprendí cuando tuve la suerte de pasar unos años trabajando en las plantas de neurología de varios hospitales. Tras examinar a pacientes con disfunciones en distintas partes del cerebro, y conversar con ellos, se me hizo evidente lo difícil que resulta que desaparezca la conciencia: siempre persiste de una forma u otra, excepto cuando se produce un coma o un estado vegetativo debido a una lesión cortical extensa que provoca una disfunción en todo el conjunto del cerebro. Este tipo de lesión puede tener varios orígenes: no haberse puesto un casco cuando más falta hacía, un coágulo o una hemorragia debida a la rotura de una arteria cerebral, la extirpación quirúrgica de un tumor situado en un mal sitio o una sobredosis de droga. En los demás casos, puede que cambie la personalidad, puede que se pierdan para siempre determinadas habilidades concretas, puede que cambie incluso la realidad personal del individuo; pero la conciencia sigue acompañándolo. Está claro que el santo grial de la ciencia es encontrar la conciencia en el cerebro, pero créame si le digo que, si hubiera «algo» que encontrar, a estas alturas ya se habría descubierto.

Durante los últimos dos mil años de la historia humana, los estudiosos han tratado de encontrar la fuente, ese «algo» —una esencia espiritual, una

glándula en la frente, un alma inmortal, una región del cerebro...— teóricamente responsable de cosas tales como el lenguaje, la memoria, la atención y la conciencia. Aunque de hecho nadie sabe cómo funciona realmente ninguna de esas capacidades que tanto apreciamos, sí sabemos qué partes del cerebro gestionan el lenguaje, la memoria y la atención; en cambio, cuando tratamos de encontrar las partes del cerebro primordialmente responsables de la conciencia, empezamos a vacilar, a sentirnos frustrados, porque parece que es imposible encontrar ese lugar. La clínica neurológica nos dice una y otra vez que debemos intentar enfocar el problema de manera distinta.

Siempre suele señalarse el hecho de que hay lesiones del tallo cerebral que tienen efectos devastadores en la conciencia, efectos tan importantes que la gente entra en coma y a menudo jamás se recupera.¹ Pero esta es una circunstancia distinta. Si desconectamos la batería de nuestro automóvil, nunca podremos verlo funcionar. No podemos arrancarlo y ver lo que es capaz de hacer. Algo parecido sucede con las lesiones del tallo cerebral: la parte principal del cerebro nunca llega a arrancar. En esos casos extremos no hay nada que observar; ni se puede aprender nada de ellos con respecto al tema de la conciencia.

Sin embargo, armados con nuestros conceptos de módulos y capas, ahora podemos abordar el desconcertante problema de la persistencia de la conciencia frente a toda una serie de lesiones devastadoras. Hemos de descubrir una forma de observar cómo se comportan la gran mayoría de los humanos cuando su cerebro no funciona bien. Necesitamos entender por qué la conciencia persiste.

En términos generales, cabe atribuir esta persistencia a la multitud de módulos que siguen contribuyendo a nuestras experiencias cotidianas pese a las posibles lesiones u otras disfunciones. Los cerebros multimodulares tienen a su disposición y recurren a una enorme cantidad de vías hacia la experiencia consciente. Si se destruye una ruta, otra puede proporcionar una vía

alternativa. Para que desaparezca la conciencia tienen que desactivarse todos los módulos que conducen a un estado consciente; en tanto que esto no suceda, los módulos intactos seguirán transmitiendo información de una capa a otra, induciendo una percepción subjetiva de la experiencia. Puede que el contenido de esa experiencia consciente sea muy distinto de lo normal, pero la conciencia permanece. En nuestra visita a la clínica neuropsicológica veremos cómo diversos ataques a nuestro cerebro afectan a la conciencia y nos proporcionan información acerca de la organización de nuestro cerebro. Resulta que las interminables fluctuaciones de nuestra vida cognitiva, gestionadas por nuestra corteza cerebral, navegan por un mar de estados emocionales que nuestro cerebro subcortical reajusta constantemente.

Una visita a la clínica

El primer paciente al que vamos a conocer podría ser el abuelo de cualquiera. Me estrecha la mano en señal de reconocimiento, pero se muestra confuso en cuanto a mi identidad: no recuerda que me conoce desde hace un par de días. Sufre el tipo de demencia más común, la enfermedad de Alzheimer, asociada a la producción y acumulación de amiloide- β en el cerebro. Eso significa que tiene un daño neuronal grave en toda su estructura cerebral. Aunque durante los últimos veinte años más o menos se ha considerado que el amiloide era la «causa» de la enfermedad de Alzheimer, hay algunas evidencias recientes que contradicen esa hipótesis, mientras que otras todavía se están contrastando.² En cualquier caso, la enfermedad supone la destrucción lenta y progresiva del cerebro, empezando especialmente por la pérdida de neuronas en la corteza entorrinal y el hipocampo, lo que provoca una pérdida de la memoria a corto plazo. La enfermedad puede debilitar al abuelo hasta el punto de transformar por completo su personalidad, pasando de ser una persona despierta y afectuosa a convertirse en una apática máscara de su antiguo yo; sin embargo, aunque puede que no me reconozca, todavía es consciente de las

convenciones sociales y me estrecha la mano. Puede que divague, pero seguirá sintiendo miedo cuando esté confuso y perdido, e ira cuando se sienta frustrado. Su experiencia consciente del mundo le llega a través de cualquier circuito neuronal operativo que siga funcionando, y, cuando pierde una función, esta se vuelve más restringida. Los contenidos de esa experiencia consciente probablemente son extraños, muy distintos de los de un cerebro normal o de su antiguo yo. Como resultado, también se produce un comportamiento extraño.

Es posible que la versión apática del abuelo anteriormente jovial, por ejemplo, todavía siga describiéndose a sí misma como su versión anterior, en la que era «el alma de la fiesta». Los cuidadores y familiares suelen atribuir esta incongruencia en la autoidentificación del paciente a la desorientación inherente a la enfermedad. Pero cuando los amigos y parientes describen la personalidad de un ser querido antes de la enfermedad, esta resulta ser llamativamente similar a la autodescripción proporcionada por el propio individuo una vez enfermo,³ lo cual sugiere que las falsas creencias del abuelo en relación con sus actuales rasgos de personalidad probablemente se deben a la incapacidad de actualizar esas mismas creencias. La demencia ha dejado al abuelo con una imagen obsoleta de sí mismo. Mientras su corazón siga latiendo, la conciencia, aunque con un contenido alterado, sobrevivirá a la masacre de la degeneración de su cerebro.

Nuestro próximo paciente es un enfermo conocido como «señor B». Este paciente tiene un problema distinto: cree que es una persona de especial interés para el FBI, que lo mantiene vigilado cada momento del día. No solo eso: además, los agentes del FBI filman y emiten su vida por televisión en un programa titulado *El show del señor B*. Comprensiblemente perturbado por ello, el señor B intenta evitar situaciones embarazosas adaptando su comportamiento. Se pone un traje de baño cada vez que se ducha, y se cambia de ropa bajo las sábanas. Evita las situaciones sociales, sabiendo que todas las personas con las que se encuentra son actores que intentan provocar una

acción dramática para hacer que *El show del señor B* resulte más intrigante. Resulta difícil imaginar cómo sería vivir en el mundo del señor B. No obstante, si se analiza detenidamente, el caso de este paciente puede revelar una situación en la que una corteza cerebral completamente normal y racional está tratando de dar sentido a ciertas anormalidades que se producen en otra región del cerebro, la llamada subcorteza.

El señor B sufre de esquizofrenia crónica. Entre los factores de riesgo de esta enfermedad se incluyen la vulnerabilidad genética y las interacciones genético-ambientales, mientras que los factores ambientales que incrementan el riesgo son, entre otros, crecer en áreas urbanizadas,⁴ ser un inmigrante en un país extraño,⁵ especialmente cuando se está socialmente aislado —como cuando se vive en un entorno integrado exclusivamente por unos pocos miembros del mismo grupo—,⁶ y consumir cannabis.⁷ Da igual qué evidencias se proporcionen al señor B para combatir sus falsas creencias: él seguirá convencido de que está siendo observado constantemente por millones de personas. Uno de los síntomas más destacados de la esquizofrenia es la percepción de que determinados estímulos concretos, identificados como carentes de importancia cuando no se sufre un episodio delirante, pasan a adquirir una extrema trascendencia personal:⁸ el tipo que levanta la vista de su periódico te está observando deliberadamente; alguien colocó la piedra en el camino para hacerte daño... Esta alteración en la percepción de la relevancia, es decir, en relación con qué es más importante y qué es objeto de mayor atención, constituye una característica tan clásica de los trastornos del espectro de la esquizofrenia que en el ámbito médico existe un creciente movimiento en favor de que se abandone la etiqueta de «esquizofrenia» y, en su lugar, se reclasifique este tipo de trastornos como «síndrome de relevancia».⁹

Un impulso sensorial adquiere relevancia cuando la señal neuronal que provoca se ve potenciada por delante de otras, lo que hace que la atención pase a centrarse en él. Shitij Kapur, psiquiatra, neurocientífico y profesor del

King's College de Londres, nos explica la diferencia entre alucinaciones y delirios: «Las alucinaciones reflejan una experiencia directa de la relevancia aberrante de representaciones internas», mientras que los delirios (falsas creencias) son el resultado de «un esfuerzo cognitivo del paciente por dar sentido a esas experiencias aberrantemente relevantes». ¹⁰ La cantidad de dopamina —un neurotransmisor— del cerebro afecta al proceso de identificación y expresión de la relevancia. Durante un estado psicótico agudo, la esquizofrenia se asocia a un aumento de la síntesis de dopamina, la liberación de dopamina y las concentraciones de dopamina sináptica en estado de reposo. ¹¹ Kapur sugiere que en la psicosis se produce una disfunción en la regulación de la dopamina que provoca la activación anormal del sistema dopaminérgico, lo que a su vez genera niveles aberrantes de este neurotransmisor y, por ende, la asignación aberrante de relevancia motivacional a objetos, personas y acciones. ¹² La investigación respalda esta afirmación. ¹³ La relevancia alterada de los estímulos sensoriales se traduce en una experiencia consciente con contenidos muy distintos de los que normalmente habría, pero esos contenidos son los que constituyen la realidad del señor B y proporcionan las experiencias a las que su cognición tiene que dar sentido. Si consideramos los contenidos de la experiencia consciente del señor B —sus alucinaciones—, sus esfuerzos por dar sentido a sus delirios ya no resultan tan extravagantes: son solo explicaciones posibles —aunque no probables— de lo que está experimentando. Teniendo esto en mente, el comportamiento que se deriva de su conclusión cognitiva parece más racional. Y pese a padecer esta función cerebral alterada, el señor B sigue siendo consciente de su existencia.

El inconsciente andante

No obstante, también pueden surgir comportamientos extraños en un cerebro completamente intacto y funcional cuando solo una parte de este está despierta.

En un cerebro estratificado tienen lugar numerosas actividades simultáneas que se coordinan sincrónicamente. Pero ¿qué ocurre si esa sincronización se pierde, si cada capa sigue funcionando pero a su propio ritmo? Nuestro próximo paciente es el señor A, quizá la más inquietante de nuestras visitas.

El señor A, descrito por familiares y amigos como un afectuoso hombre de familia, se despertó una noche en la cama oyendo el ladrido de sus perros y voces de personas extrañas. Tras bajar corriendo las escaleras de casa, se encontró con varios agentes de policía que le apuntaban con sus armas.¹⁴ Aturdido y confuso, fue esposado e introducido en el asiento trasero de un coche patrulla, temblando de miedo mientras intentaba evaluar la situación escuchando las conversaciones del personal de emergencia a través de la ventanilla. Pudo oír que su esposa había sido gravemente herida, y pensó que la policía estaba buscando al responsable. No supo hasta más tarde que en realidad ya lo habían encontrado, y que era él.

Presa de un tremendo pánico, el señor A solo recordaba que se había quedado dormido en su cama unas horas antes. La policía aclaró la trágica situación. El señor A había asesinado brutalmente a su esposa durante lo que luego se determinó que había sido un episodio de sonambulismo. Durante ese episodio, se había levantado de la cama y había bajado a arreglar el filtro de la piscina, como su esposa le había pedido que hiciera durante la cena. Ella debió de despertarse, y bajó para convencerle de que volviera a la cama. Al ver interrumpida su concentración en el arreglo del mecanismo, se volvió violento y le asestó cuarenta y cinco puñaladas; luego guardó las herramientas en el garaje, volvió a buscar a su mujer —que todavía seguía con vida— y la empujó a la piscina, donde se ahogó. Después volvió a acostarse. Al oír gritos y ladridos en la casa de al lado, su vecino se asomó por encima de la valla y vio cómo el señor A, aparentemente «confuso», arrojaba un cuerpo a la piscina; entonces llamó a la policía.

La idea de que alguien pueda matar a su esposa, a la que ama, durante un episodio de sonambulismo resulta aparentemente inexplicable. Sin embargo,

en ausencia de un motivo identificable, y dado que no había hecho el menor intento de ocultar el cuerpo o el arma ni recordaba en absoluto los hechos, el jurado se convenció de que los actos del señor A se habían producido de forma involuntaria y sin que él fuera consciente de ello. Si eso es cierto, ¿qué sucedió exactamente en la mente y el cerebro del señor A mientras cometía aquella atrocidad?

El sonambulismo es una «parasomnia», esto es, un comportamiento anormal que se produce durante el sueño. Con los años, los expertos han identificado dos fases principales del sueño mediante el registro de las ondas cerebrales: el llamado «sueño de movimientos oculares rápidos» (o REM, por sus siglas en inglés) y el «sueño sin movimientos oculares rápidos» (denominado no REM). Por regla general, el sonambulismo ocurre tras un despertar espontáneo abrupto e incompleto del sueño no REM producido en las primeras dos horas de la noche, convirtiendo al sujeto en un durmiente móvil. Intentar despertar a un sonámbulo a menudo resulta infructuoso, y además puede ser peligroso, ya que este puede sentirse amenazado por el contacto físico y responder de manera violenta. Normalmente el sueño no REM da paso al sueño REM, durante el cual se produce una pérdida de tono muscular, lo que impide el comportamiento motor durante esa fase. La mayoría de los episodios de sonambulismo tienden a ser relativamente inofensivos y suelen ser fuente de una buena historia que contar por parte de los testigos, que a menudo comienza con: «¡No vas a creerte lo que hiciste anoche!». Y si el sonámbulo eres tú, desde luego que no te lo crees, puesto que no tendrás el menor recuerdo de tus correrías nocturnas.

La mayoría de los comportamientos parasómnicos parecen irracionales y resultan desconcertantes de observar. Al sonámbulo le puede dar por pasar la aspiradora o por barrer el patio en mitad de la noche, totalmente ajeno a su entorno. En ocasiones, los sonámbulos se entregan a actividades más complejas y potencialmente peligrosas, como cortar el césped, reparar una motocicleta o conducir un coche. Este tipo de comportamientos complejos

hacen difícil creer que en ese momento el sonámbulo no tiene una percepción consciente de sus actos. Pocas veces esos comportamientos complejos se vuelven violentos. Cuando intervienen las fuerzas de la ley, determinar si el comportamiento fue intencionado pasa a adquirir una importancia fundamental, lo que intensifica el debate acerca de si estas personas son conscientes o no de las actividades que realizan durante sus períodos de sonambulismo.

Las técnicas de neuroimagen y de electroencefalografía nos han permitido tener una idea más clara de lo que sucede en el cerebro durante el sueño no REM,¹⁵ el sonambulismo¹⁶ y los denominados «despertares confusos».¹⁷ Parece que en tales situaciones el cerebro está medio dormido y medio despierto: el cerebelo y el tallo cerebral están activos, mientras que el cerebro y la corteza cerebral tienen una actividad mínima. Las vías neuronales involucradas en el control del comportamiento motor complejo y la generación de emociones se mantienen en funcionamiento, mientras que las que se proyectan hacia el lóbulo frontal, relacionadas con la planificación, la atención, el juicio, el reconocimiento facial emocional y la regulación emocional, se desconectan. Los sonámbulos no recuerdan sus correrías, ni tampoco les despiertan el ruido o los gritos, puesto que las partes de la corteza que contribuyen al procesamiento sensorial y a la formación de nuevos recuerdos dormitan, temporalmente desconectadas, y, por lo tanto, no envían señal alguna al flujo de la conciencia.

Es probable que el señor A experimentara conscientemente algunos aspectos del episodio, pero de forma muy distinta de lo que lo haría su yo despierto. Teniendo presente el concepto de cerebro estratificado, podemos predecir que ciertos módulos de producción de conciencia de «nivel inferior» estaban activos, permitiéndole orientarse y coordinar sus movimientos con destreza y sentir emociones, mientras que otros, de «nivel superior», permanecían dormidos y en silencio, impidiéndole comprender la situación, reconocer a su esposa, oír sus gritos o recordar los hechos. En el conjunto del sistema había determinadas regiones aisladas y desconectadas, y solo ciertos

módulos contribuían a su comportamiento y a su experiencia consciente. Por desgracia, la corteza estaba dormida, inmovilizada y sin aportar nada. Cuando el señor A salió de su ciclo de sueño, esos módulos silenciosos despertaron a una realidad de pesadilla. Durante aquel horrendo incidente, las regiones despiertas de su cerebro intacto, libres de las trabas del procesamiento de los módulos de control cognitivo de su corteza, ahora dormida, generaron un comportamiento que se apartó de manera considerable de la conducta habitual de aquel hombre normalmente compasivo y no violento. El hecho de que aquellos actos fueran absolutamente contrarios a los rasgos e ideales de la personalidad del señor A fue exactamente la razón por la que el jurado emitió un veredicto de «inocente».

Inmóvil, pero consciente

En contraste con el caso anterior, una de las lesiones cerebrales más espeluznantes es la producida en la parte ventral del puente troncoencefálico. La pérdida de estas neuronas, que conectan el cerebelo con la corteza, deja a quien la sufre sin poder moverse, pero totalmente consciente. Un ejemplo es el célebre caso de Jean-Dominique Bauby, redactor jefe de la revista *Elle*, que sufrió un derrame cerebral a los cuarenta y tres años de edad. Tras despertarse varias semanas después de un coma, completamente consciente y sin ninguna pérdida cognitiva, se encontró con que no podía mover ninguna parte de su cuerpo excepto el párpado izquierdo.¹⁸ Eso significaba asimismo que tampoco podía hablar y, por lo tanto, no podía decirle a nadie que estaba consciente. Tuvo que esperar a que alguien advirtiera que parecía mover el párpado de manera voluntaria. Esta dolencia se conoce como «síndrome de enclaustramiento». Los más afortunados —si es que se les puede llamar así— pueden parpadear o mover los ojos a voluntad, aunque dicho movimiento resulta poco perceptible y agotador. Pero así se comunican; los más desafortunados no pueden hacerlo.

En muchos casos han transcurrido meses o incluso años antes de que algún cuidador reconociera que el paciente estaba consciente, sufriendo mientras tanto intervenciones médicas sin anestesia y oyendo conversaciones sobre su propia suerte en las que no podía participar. Cuando se advirtió que Bauby estaba consciente, este pudo aprovechar su capacidad de pestañear. Gracias a ello escribió un libro en el que describía su experiencia consciente mientras yacía inmóvil. Tendido en su lecho, construía y memorizaba oraciones. Luego, durante cuatro horas al día, un amanuense se sentaba pacientemente a su cabecera e iba leyendo un alfabeto francés ordenado por frecuencias; Bauby pestañeaba cuando se pronunciaba la letra correcta. Doscientos mil parpadeos después se completó *La escafandra y la mariposa*. En el prólogo, hablando en tercera persona, Bauby describe la situación en la que se encontró al despertar del coma: «Paralizado de pies a cabeza, el paciente, con la mente intacta, está confinado dentro de su propio cuerpo, sin poder hablar ni moverse. En mi caso, el pestañeo del párpado izquierdo es mi único medio de comunicación». ¹⁹ Explica que se siente agarrotado y que es capaz de experimentar dolor, pero luego añade:

Mi mente alza el vuelo como una mariposa. Hay mucho que hacer. Puedes vagar por el espacio o el tiempo, viajar a la Tierra del Fuego o a la corte del rey Midas.

Puedes ir a ver a la mujer que amas, deslizarte a su lado y acariciar su rostro mientras aún duerme. Puedes construir castillos en España, robar el Vellocoino de Oro, descubrir la Atlántida, realizar los sueños de tu infancia y tus ambiciones adultas. ²⁰

El caso de Bauby es un buen ejemplo de la inagotable capacidad de adaptación humana. De hecho, la capacidad de adaptación parece ser una constante en estos pacientes, ya que el 75 % de ellos raras veces o nunca han tenido pensamientos suicidas. ²¹ Incluso con esta devastadora lesión en una parte del tallo cerebral, la conciencia permanece, acompañada de toda la consiguiente gama de sentimientos tanto sobre el presente como sobre las experiencias pasadas.

Cuando hay módulos y capas del cerebro que están dañados o funcionan

mal, pueden producirse comportamientos extraños. Desde el daño y la alteración cortical generalizados propios de la enfermedad de Alzheimer hasta los trastornos específicos asociados a las lesiones del tallo cerebral, empieza a emerger una idea: es necesario entender el funcionamiento tanto de la corteza como de la subcorteza para capturar los momentos siempre cambiantes de la experiencia consciente. ¿Es posible que todos esos fugaces pensamientos conscientes tengan como base unos pocos estados emocionales concretos que doten a dichos pensamientos de una cualidad subjetiva? ¿Es posible que todo esto encaje en la arquitectura estratificada del cerebro, con el sistema cerebral evolutivamente más antiguo —todavía configurado para indicarle al organismo si debe luchar o correr, buscar pareja o comer— operando fuera del control directo de las capas cognitivas? ¿Acaso el modelo de la arquitectura estratificada nos proporcionará los medios para comprender cómo estamos organizados para ser conscientes?

El motor subcortical de las emociones

Durante largo tiempo se ha creído que la corteza cerebral es la responsable de todas las formas de conciencia, y que sin ella no solo seríamos simplemente inconscientes, sino que nos veríamos privados de la capacidad de ser conscientes a todos los niveles, es decir, que seríamos seres sin conciencia en estado vegetativo.²² Sin embargo, es posible que la corteza sea simplemente una colección de extensiones (o aplicaciones) para potenciar las experiencias conscientes. Es cierto que esta nos proporciona varios conjuntos de habilidades mentales dinámicas —es decir, que pueden cambiar y que están constantemente activas—, pero podría no ser esencial para proporcionarnos una subjetividad pura y dura. Bajo la capucha cortical hay varias redes subcorticales que resultan cruciales para mantener la conciencia: son precisamente las lesiones de estas regiones subcorticales las que pueden provocar un coma, el estado en que una persona o animal deja de responder a

los estímulos y parece inconsciente a ojos de un observador externo.²³ Ni siquiera una corteza cerebral completamente funcional puede rescatar los restos del naufragio de ciertos tipos de lesiones subcorticales.

En el pasado, el reto de trazar la línea divisoria entre un estado consciente y un estado inconsciente ha sido en gran parte de índole semántica. El término *conciencia* carece de objetividad, puesto que resulta difícil definir la percepción subjetiva de la propia existencia; esa es la razón principal de que las características definitorias de la conciencia sean siempre objeto de un acalorado debate. Pero cuando uno entra en el ámbito clínico la necesidad de identificar los estados conscientes resulta de la máxima urgencia, y el problema deja de ser meramente semántico para convertirse también en una cuestión ética. Negar la medicación analgésica a un paciente aparentemente inconsciente que en realidad está consciente sin que lo sepamos es tortura. Pese a la ambigüedad que rodea el término, hay pruebas convincentes que sugieren que la corteza cerebral no es necesaria para suscitar algunas formas de conciencia. Las capacidades de los sistemas subcorticales parecen ser lo bastante competentes por sí solas como para proporcionar una experiencia subjetiva.

Estas evidencias provienen de conclusiones extraídas de la clínica pediátrica. Lamentablemente, algunos niños nacen con anencefalia (sin corteza cerebral, debido a causas genéticas o a deficiencias de desarrollo) o con hidranencefalia (con una corteza cerebral mínima, a menudo resultado de un traumatismo o enfermedad fetal). El neurocientífico Björn Merker se interesó en el campo de la subcorteza cerebral ya desde los comienzos de su carrera. Frustrado por la limitada información sobre el tema y el reducido número de estudios de casos de niños con hidranencefalia, se unió a un grupo de Internet integrado por padres y cuidadores de todo el mundo a fin de saber más sobre ellos y su dolencia. Conoció a varias familias y pasó una semana con ellas en Disney World. En el transcurso de aquella semana observó que los niños «no solo están despiertos y a menudo alerta, sino que muestran receptividad a su

entorno en forma de reacciones emocionales u orientadoras a los acontecimientos de este. [...] Expresan placer sonriendo y soltando risotadas, y aversión “armando bulla”, arqueando la espalda y llorando (en muchas gradaciones), al tiempo que su rostro se anima por esos estados emocionales. Un adulto familiarizado puede utilizar esa receptividad para construir secuencias de juego que progresen de manera predecible desde la sonrisa, pasando por las risitas ahogadas, hasta la carcajada acompañado de un gran entusiasmo por parte del niño». ²⁴ Aun careciendo de corteza cerebral o de la cognición que esta proporciona, esos niños sentían emociones, tenían experiencia subjetiva y poseían conciencia. Nadie los confundiría nunca con niños dotados de corteza cerebral, pero son conscientes, y su respuesta emocional a los estímulos es apropiada.

Con los años, Merker ha llegado a la conclusión de que es el mesencéfalo el que sustenta la capacidad básica de la experiencia subjetiva consciente. Es cierto que la corteza elabora los contenidos de la experiencia, pero esa capacidad en sí misma surge de las estructuras del mesencéfalo. Las implicaciones éticas de esta afirmación resultan evidentes. Merker explica que los padres suelen encontrarse con profesionales médicos que se sorprenden cuando les piden medicamentos analgésicos para estos niños cuando tienen que someterse a procedimientos invasivos.

El principal argumento en contra de la idea de que esos niños experimentan el mundo a través de sus estructuras subcorticales se basa en el hecho de que en casi todos ellos hay alguna parte de la corteza cerebral que se ha preservado. Sin embargo, aun cuando esas regiones corticales intactas, siempre muy limitadas (y de cuestionable funcionalidad), varían ampliamente de un niño a otro, su comportamiento es bastante homogéneo y no guarda proporción alguna con el tejido presente. Así, por ejemplo, mientras que el tejido cortical auditivo rara vez se conserva, en cambio se mantiene la audición; y aunque es frecuente que se preserve una parte de la corteza visual, la visión tiende a verse comprometida.

La teoría de Merker se ve reforzada por las investigaciones sobre la vida emocional de los animales. El neurocientífico de origen estonio Jaak Panksepp, que dedicó medio siglo a estudiar la naturaleza de las emociones en los animales, diferenciaba dos tipos de conciencia: por una parte, la — evolutivamente antigua— conciencia afectiva (consciente de las emociones en bruto) y, por otra, una relativa recién llegada a escala evolutiva, la conciencia cognitiva (la que permite pensar en esas emociones). En una de sus conferencias, explicó cómo era la práctica de laboratorio que exigía a sus alumnos a final de curso. Preparaba dos ratas para cada estudiante. Una de ellas había sido decorticada, dejándole solo el tejido subcortical; a la otra se la había sometido a una cirugía ficticia, es decir, que se le había practicado una operación quirúrgica en la que en realidad no se le había extirpado nada del cerebro. Los estudiantes debían examinar a su pareja de ratas durante dos horas, haciéndoles realizar una amplia gama de tareas que habían estudiado en clase. Al finalizar el tiempo estipulado, tenían que adivinar cuál de las dos ratas había perdido toda la corteza cerebral y explicar el porqué de su elección. Doce de los dieciséis integrantes de la clase declararon que las ratas decorticadas eran las neurológicamente normales.

Lo que los estudiantes observaron en aquellas ratas eran comportamientos motivados, como buscar comida, aparearse, luchar o huir cuando eran atacadas, y jugar simulando peleas con otras ratas;²⁵ para aquellos estudiantes, los comportamientos observados eran lo suficientemente característicos de las ratas como para declararlos normales. Si la corteza fuera la única responsable de mediar la conciencia, su extirpación debería haber hecho que aquellas ratas no respondieran a los juegos de las otras ratas ni a todo lo demás. Pero la extirpación de la corteza cerebral no eliminó sus competencias y respuestas básicas, lo que significa que los mecanismos superiores del tallo encefálico bastaban por sí solos para sustentar muchos de sus comportamientos, incluyendo sus sentimientos emocionales y motivacionales, que analizaremos a continuación.

Sentimientos vinculados a la conciencia

Todos estos estudios sobre ratas decorticadas y niños con hidranencefalia sugieren que las estructuras subcorticales pueden transformar el estímulo neuronal en bruto en algo parecido a las emociones básicas. Las áreas subcorticales del cerebro tienen su propia dinámica, surgieron muy pronto en el proceso evolutivo y son anatómica, neuroquímica y funcionalmente homólogas en todos los mamíferos que se han estudiado.²⁶ Panksepp argumentaba que compartimos con otros animales las áreas cerebrales que producen las emociones que sentimos, y que la evolución ha seleccionado dichas áreas por su capacidad para incrementar las posibilidades de supervivencia. ¿De qué modo? Las emociones actúan como un sistema de recompensa y castigo interno que informa a un animal de cómo le está yendo en la lucha por la supervivencia. Las emociones positivas incitan al animal, mientras que las negativas, dependiendo de su intensidad, denotan toda una gama de situaciones que van desde las meramente inciertas hasta las potencialmente desastrosas. En consecuencia, esos sentidos internos proporcionan una forma de evaluar el entorno exterior y son potentes motores del comportamiento, a pesar de su relativa simplicidad en el nivel de la conciencia.

David Anderson y Ralph Adolphs, investigadores del Instituto de Tecnología de California, son de la misma opinión que Panksepp.²⁷ Sostienen que una emoción es un estado inconsciente del sistema nervioso central desencadenado por un estímulo específico, ya sea externo, como un depredador, o interno, como el recuerdo de uno. Cuando se activa, el circuito neuronal que codifica ese estado pone en marcha múltiples procesos paralelos que a su vez generan una respuesta conductual, además de sentimientos, cambios cognitivos y respuestas somáticas como, por ejemplo, la aceleración de los latidos cardiacos y la sequedad de boca. Es decir, que podemos sentir sin que la cognición entre en juego y dé cuenta de ese sentimiento.

Para Panksepp, hay siete sentimientos emocionales y motivacionales

primarios que parecen ser rasgos comunes de la conciencia animal y humana tanto a nivel conductual como neuronal; son: búsqueda, temor, ira, deseo, cuidado, aflicción y juego. Estos sentimientos, que en gran parte pueden atribuirse a las funciones del sistema límbico subcortical, llevan a los animales —incluidos los humanos— a comportarse de formas que favorecen la búsqueda de alimento, refugio y pareja; la evitación del daño; la protección de sí mismos y de su parentela, y el establecimiento de relaciones sociales con familiares y amigos. Si consideramos que la conciencia es un sentimiento subjetivo sobre algo, debemos considerar que las emociones constituyen uno de sus componentes fundamentales.

Panksepp concluía que las emociones constituyen una herramienta vital tan fructífera que se codificaron en el genoma en una forma primaria, se han conservado en todas las especies de mamíferos, y solo en una fase posterior del proceso evolutivo se revistieron de los mecanismos de aprendizaje y las cogniciones de orden superior proporcionados por una extensión adicional: la corteza.²⁸ Si esos sentimientos ya existían antes que los tejidos corticales, entonces es que el especial cableado de esas redes subcorticales debe de poseer por sí solo todo lo necesario para producir los sentimientos que acompañan a la experiencia consciente. Si entendemos el funcionamiento del sistema estratificado de las redes subcorticales, quizá podamos apreciar mejor la forma más primitiva de conciencia. Los sentimientos emocionales y motivacionales, así como las conductas que generan en los animales, pueden enseñarnos mucho acerca de cómo los sistemas modulares favorecen la conciencia, y tal vez puedan señalararnos también qué aspectos de la conciencia humana son exclusivos de esta.²⁹

Joseph LeDoux, de la Universidad de Nueva York, que se ha dedicado a dilucidar minuciosamente el funcionamiento de lo que antes denominaba los «circuitos del miedo» pero ahora prefiere llamar «circuitos de amenaza», adopta una perspectiva distinta. Para él, existen dos importantes motivos de inquietud: el primero es que todavía no haya una definición consensuada de

emoción; el segundo, que haya quienes cuestionan la idea de que existen emociones básicas comunes. De ser así, ¿cómo podemos diferenciar con seguridad la emoción de otros estados psíquicos o comparar emociones entre distintas especies? Escribe LeDoux: «La respuesta corta es que lo simulamos. Las introspecciones de experiencias subjetivas personales nos dicen que algunos estados mentales tienen un cierto “sentimiento” asociado a ellos, mientras que otros no». Esto le ha llevado a interesarse por la afirmación de que un comportamiento similar en los animales indica también una experiencia similar.³⁰ Desde su perspectiva, la corteza es necesaria para los sentimientos afectivos. Él cree que los circuitos subcorticales producen un comportamiento emocional y unas respuestas fisiológicas determinados, pero que estos solo contribuyen de manera indirecta a los sentimientos subjetivos. Considera que para la producción de sentimientos subjetivos se requiere un paso cognitivo adicional, y que dicho paso lo proporcionan los circuitos corticales superiores que leen e interpretan el comportamiento emocional. No es el único que piensa así. De hecho, la mayoría de los investigadores del ámbito de las emociones darían su aprobación a estas teorías sobre la «lectura» cognitiva. LeDoux sugiere que los sentimientos conscientes constituyen un proceso en dos etapas y que se producen cuando las diversas partes de la corteza prefrontal que sustentan la memoria de trabajo leen una respuesta fisiológica.

Mientras se libra esta batalla en torno a las emociones, nosotros podemos mantenernos al margen, ya que la noción de una arquitectura cerebral estratificada se adapta a cualquiera de los dos escenarios. Lo importante es que tanto la subcorteza como la corteza contribuyen a la plena experiencia consciente. Desde una determinada perspectiva, los niños con hidranencefalia experimentan emociones que parecen idénticas a las de los niños con la corteza intacta. Dado que su comportamiento manifiesto es similar, nos apresuramos a proyectar también en ellos el complemento íntegro de una experiencia metaautoconsciente (es decir, la conciencia de que son conscientes). ¿Realmente tienen conciencia de sí mismos? Sin una corteza que

proporcione las funciones necesarias para la cognición, no pueden saber que son autoconscientes. Como mínimo, para tener plena conciencia de que uno está teniendo una experiencia consciente, tienen que funcionar ambas capas.

La conciencia potenciada por la corteza

Así pues, dado que los circuitos subcorticales contienen los ingredientes esenciales para la conciencia, ¿estamos sobrevalorando el papel de la corteza? ¡Por supuesto que no! Lo importante aquí es no infravalorar el de la subcorteza. Entendiendo la contribución del procesamiento subcortical a la conciencia nos hallamos mejor equipados para comprender por qué es tan difícil librarse de este flujo incesante de «sentimientos sobre sentimientos». Está claro que la corteza cerebral desempeña la función de proporcionar los contenidos de la conciencia, dadas las evidencias de que las lesiones cerebrales en esta zona suelen coincidir con cambios conductuales específicos. ¿Y cuál es exactamente el papel de la corteza cerebral en la producción de la conciencia? La corteza aumenta el número de formas en que podemos experimentar el mundo, lo que permite una enorme variedad de posibles experiencias y respuestas conscientes.

El tipo peculiar de corteza que posee cada especie le proporciona sus propios contenidos específicos de experiencia consciente. Parte del contenido de la conciencia humana es el lenguaje. Solo los humanos han inventado una serie de pequeños e ingeniosos símbolos que, en una determinada combinación, pueden proporcionar a otra persona una representación mental concreta de alguna idea abstracta. No solo tenemos la capacidad de aprender el lenguaje, sino que también estamos biológicamente preparados para su adquisición.³¹ Como veíamos en el Capítulo 4, disponemos de regiones cerebrales enteras dedicadas a los diversos aspectos del aprendizaje, la comprensión y la producción del lenguaje. Otra visita a la clínica nos mostrará que una lesión en una de estas regiones destruirá nuestra capacidad de

entender las palabras, pero seguirá permitiéndonos producir frases gramaticalmente correctas, con la prosodia y la entonación apropiadas, aunque, eso sí, sin el menor sentido. Una lesión en un área distinta nos permitirá entender las frases, pero no construirlas. Si sufrimos una lesión en otra no podremos pronunciar sustantivos, pero seguiremos siendo capaces de reconocerlos y entenderlos. Cualquiera de estas lesiones se traducirá en una experiencia consciente distinta, pero ninguna de ellas destruirá la conciencia en sí.

Si bien el lenguaje se suma a nuestras experiencias conscientes, seguiríamos siendo conscientes sin él, aunque muchas de nuestras experiencias serían notoriamente distintas. Consideremos el caso del «niño salvaje» francés Víctor de Aveyron, inmortalizado en la película de François Truffaut *El pequeño salvaje*, estrenada en 1970. Víctor, descubierto cuando tenía doce años de edad, había pasado toda su infancia viviendo solo en el bosque, nunca había estado en contacto con el lenguaje humano y, en consecuencia, no había aprendido a hablar. Era claramente consciente y tenía experiencias conscientes, pero con contenidos distintos de los que habría tenido de haber aprendido a hablar. Cuando la funcionalidad de un módulo no se desarrolla, entran en juego otros módulos para brindarnos una experiencia alternativa.

Existe un amplio debate en torno a la cuestión de si las estructuras subcorticales son o no la principal fuerza motriz de la conciencia³² o de si esta última está mediada primordialmente o no por la corteza cerebral.³³ Pero si se piensa en el funcionamiento del cerebro, cabe deducir que quizá no haya una jerarquía modular específica que permita que la conciencia se manifieste de una forma u otra. Cada módulo concreto funciona de manera relativamente independiente de los demás, y, lejos de ser una cola pulcramente ordenada de procesamientos modulares, puede que los contenidos de nuestra experiencia sean el resultado de una especie de competencia: en un momento en concreto algunos procesos se adueñan de nuestro paisaje consciente, mientras que otros no. Según esta hipótesis, tanto los módulos subcorticales como los corticales

tienen la capacidad de producir una forma de experiencia consciente que no requiere necesariamente la intervención de sistemas cognitivos «inferiores» o «superiores». Lejos de ello, la multitud de módulos productores de la conciencia se limita a «diversificar nuestra cartera» consciente. Para ilustrar mejor este concepto, vamos a intentar erradicar la conciencia con una barra de hierro.

Una de las lesiones cerebrales más fascinantes y famosas de la historia surgió como consecuencia de una explosión producida en los trabajos de construcción de una vía férrea, que incrustó una barra de metal ardiente en el cráneo y el lóbulo frontal izquierdo de un obrero llamado Phineas Gage. Sorprendentemente, Phineas pareció no perder la conciencia ni siquiera en el momento del accidente. Personalmente preferiría experimentar el famoso puñetazo de Muhammad Ali que dejó fuera de combate a Sonny Liston antes de que me reventara el cráneo una barra de hierro, aunque al parecer el golpe de Ali resultó más eficaz a la hora de eliminar la conciencia (al menos temporalmente). Mientras que Sonny se recuperó del puñetazo, en el caso de Phineas Gage se produjo un daño cerebral inmediato y permanente. No obstante, a pesar de perder la mitad del lóbulo frontal, Phineas pudo seguir funcionando de manera similar a como lo hacía antes del accidente; salvo en un aspecto: sus modales cambiaron drásticamente y el otrora profesional y respetuoso Phineas se convirtió en un hombre lascivo e irrespetuoso.³⁴ Se volvió menos concienzudo en su trato con el prójimo, aunque en absoluto menos consciente. Su rango de posibles experiencias conscientes disminuyó un poco, ya que al parecer sus actitudes antaño empáticas hacia sus colegas se vieron reemplazadas por experiencias de agitación y agresión. Phineas Gage sufría lo que hoy se conoce como síndrome del lóbulo frontal, por el que perdió toda la funcionalidad de la parte izquierda del lóbulo frontal. Cuando se produce una lesión en el lóbulo frontal, las personas suelen tener dificultades para regular sus emociones.³⁵ Esta pérdida de control emocional puede atribuirse al hecho de que los módulos subcorticales «ganan» con mayor

frecuencia la competición para proporcionar una experiencia consciente generalizada, dado que hay una menor competencia moduladora de los tejidos frontales. Pero, sea cual fuere la razón subyacente de la pérdida de control emocional en el síndrome del lóbulo frontal, hay una constante en todos los casos: la persona sigue siendo consciente.

Existe una enorme cantidad de casos de lesiones cerebrales que presentan un panorama similar: una lesión o disfunción en la región X del cerebro causa un cambio en el comportamiento Y, pero la conciencia casi siempre se mantiene intacta. El cerebro modular hace resiliente la conciencia gracias a la gran cantidad de posibles vías que pueden conducir a un momento consciente. Solo un cerebro organizado de ese modo puede explicar estos hechos neurológicos. Perder módulos causa pérdidas de funcionalidades concretas, pero la mente sigue produciendo un constante flujo consciente como si nada hubiera cambiado. Lo único que cambia es el contenido de ese flujo. Este hecho no solo proporciona una evidencia de que el cerebro funciona de forma modular, sino que también sugiere que cada uno de esos módulos independientes puede producir una forma única de conciencia.

La ubicuidad de la conciencia

Lo que hemos aprendido de nuestras visitas a la clínica neurológica es que las lesiones cerebrales graves en distintas ubicaciones del cerebro no bastan para erradicar la conciencia *per se*. Es posible que se pierdan ciertos contenidos de la experiencia consciente, pero no la conciencia en sí misma. Este hecho sugiere que no existe un circuito cortical específico tipo «estación central» que produzca la conciencia, sino que puede producirla cualquier parte de la corteza respaldada por el procesamiento subcortical, y que este último por sí solo es capaz de sustentar un tipo limitado de experiencia consciente. Parece, pues, que es el procesamiento de circuitos modulares locales el que proporciona los contenidos de la experiencia consciente.

Aunque estos sistemas modulares son en gran medida independientes, la comunicación entre módulos ayuda a coordinar el flujo de la conciencia. Dicha comunicación es importante con vistas a mantener cada módulo actualizado en relación con los acontecimientos personales recientes. Al igual que las noticias informan a los ciudadanos de los acontecimientos del mundo a medida que estos se producen, las conexiones entre módulos coordinan la información para garantizar que todos los módulos funcionan de manera sincronizada. Nosotros solo somos conscientes de ello cuando la comunicación llega demasiado tarde. Oír crujidos en la puerta trasera de casa en plena noche puede activar la respuesta subcortical de lucha o huida y hacer que nos precipitemos a coger el teléfono para llamar a la policía, solo para darnos cuenta al cabo de un momento, cuando entra en juego la cognición, de que, una vez más, se trata del mapache que suele revolver en nuestra basura. Debemos agradecer a nuestro sistema límbico que nos prepare tan rápidamente para una situación potencialmente peligrosa, pero en este caso no hay peligro: solo otro desastre que afrontar de buena mañana.

Esta incesante interacción entre la cognición y los sentimientos, es decir, entre los módulos corticales y subcorticales, produce lo que llamamos «conciencia». Obviamente, es una sensación muy distinta experimentar una oleada de emoción intensa que forjar un pensamiento abstracto, pero cada forma consciente es una experiencia que nos proporciona una percepción única de la realidad, y la pauta en la que estas diversas formas conscientes entran y salen de la conciencia configura nuestra propia historia vital personal. La hipótesis que mejor explica la gran variedad de formas conscientes y la ubicuidad de la conciencia en el cerebro es la de la arquitectura modular; pero hoy el reto conceptual consiste en comprender cómo cientos de módulos, si no miles, integrados en una arquitectura estratificada —cada una de cuyas capas es capaz de producir una forma de conciencia por sí sola—, nos proporcionan una experiencia vital única y uniforme en un momento concreto en el tiempo que parece fluir perfectamente hacia el momento siguiente. La idea clave aquí,

como analizaremos en el Capítulo 9, es el tiempo; una incesante secuencia en la que cada módulo tiene su momento.

Pronto veremos cómo funciona esto, pero antes de llegar a ello tenemos que afrontar una verdad evidente que a menudo pasa inadvertida. Cualquiera que sea el modelo que uno tenga de cómo se las arregla el cerebro para convertir la activación neuronal en acontecimientos mentales, necesitamos tratar de entender la distancia que separa esos dos fenómenos, uno objetivo (neuronal) y otro subjetivo (mental), y determinar si es posible siquiera salvar esa distancia. Independientemente de si pensamos que los responsables de lo que denominamos «estados conscientes» son los módulos locales o creemos que lo que subyace a dichos estados son circuitos cerebrales centrales, el problema de la distancia sigue estando ahí, y tenemos que abordarlo. No obstante, hay quien considera que es una tarea imposible.

Para llegar a esta cuestión fundamental, tenemos que echar un vistazo retrospectivo a lo que los físicos y matemáticos han pensado durante los últimos ciento cincuenta años. Al fin y al cabo, fue a ellos a quienes se les ocurrieron las que probablemente sean las más grandes ideas de la historia humana: la teoría de la relatividad y la teoría cuántica. Su pensamiento ha configurado la auténtica vanguardia de la capacidad mental humana a pesar de tener que lidiar con tremendas incógnitas y, sin embargo, los frutos de ese pensamiento han sido prácticamente ignorados por los biólogos, psicólogos y neurocientíficos, que los han desechado tildándolos de irrelevantes para el problema de la conciencia; sin embargo, yo creo que pueden ayudar a la resolución de este problema. Y lo creo por una razón: la idea que más se ha subestimado de las matemáticas y la física es la idea de «complementariedad», que afirma que una misma cosa puede tener dos tipos de descripción y de realidad. ¿Podría esta idea ayudarnos a abordar la profunda división entre mente y cerebro? ¿Podría ayudarnos a entender el denominado «vacío explicativo» (o «brecha explicativa») entre la realidad del mundo físico —ese cerebro material nuestro formado por sustancias químicas

governadas por las leyes de la física— y la realidad aparentemente inmaterial de nuestra experiencia subjetiva? Creo que sí, pero antes de seguir vamos a familiarizarnos con la física que puede contener esa clave.

PARTE III

Llega la conciencia

El concepto de complementariedad: el regalo de la física

Quienes no se sorprenden cuando se tropiezan por primera vez con la teoría cuántica posiblemente no la han entendido.

NIELS BOHR

En 1868, John Tyndall —físico, montañero, educador y profesor de la Royal Institution de Londres— dio una charla en el departamento de física y matemáticas de la Asociación Británica para el Progreso de la Ciencia. En ella expuso el siguiente dilema:

El paso de la física del cerebro a los hechos correspondientes de la conciencia resulta inconcebible. Aceptemos que un pensamiento definido y una acción molecular definida en el cerebro se producen de manera simultánea; no poseemos el órgano intelectual, ni aparentemente ningún rudimento de dicho órgano, que nos permita pasar de uno a otro por un proceso de razonamiento [...] «¿Cómo se relacionan esos procesos físicos con los hechos de la conciencia?» El abismo entre estos dos tipos de fenómenos seguiría siendo intelectualmente insalvable.¹

He aquí que, ciento cincuenta años después, no hemos logrado avanzar tanto como quisiéramos. Entendemos, hasta cierto punto, las descargas eléctricas, las agrupaciones y el flujo de moléculas, y en ocasiones incluso los correspondientes estados cerebrales, especialmente en el estudio de la visión; sin embargo, a diferencia de Tyndall, yo creo que sí tenemos un órgano que está a la altura de esta tarea. Lo que hace falta es aplicar las ideas correctas al problema de determinar cómo surge la mente a partir del cerebro. ¿Cómo concebimos esa molesta brecha entre nuestra biología y nuestra mente?

Habitualmente, ese abismo o brecha se considera un problema. Hace tan solo veinticinco años que el filósofo Joseph Levine lo denominó oficialmente «brecha explicativa»,² un concepto que luego pasó a describir en su libro *Purple Haze*:

Yo sostengo que no tenemos ni idea de cómo un objeto físico podría constituir un sujeto de experiencia, que disfruta —y no simplemente representa— una serie de estados con todo tipo de caracteres cualitativos. Cuando ahora miro mi caja de disquetes de color rojo, estoy teniendo una experiencia visual de carácter rojizo. La luz de una determinada composición se refleja en la caja de disquetes y estimula mi retina de una determinada manera. Esa estimulación retiniana genera nuevos impulsos a través del nervio óptico, provocando en última instancia varios acontecimientos neuronales en la corteza visual. ¿En qué parte de todo esto se pueden ver los acontecimientos que explican que yo tenga una experiencia rojiza? No parece haber una conexión discernible entre la descripción física y la mental y, por ende, tampoco una explicación de esta última en términos de aquella.³

Levine nos deja, pues, con un abismo insalvable entre el nivel físico de las neuronas que interactúan entre sí y el nivel aparentemente difuso de la experiencia consciente. Podemos explicar, por ejemplo, que el dolor está causado por la estimulación de las fibras C del sistema nervioso, y también por qué existe una pequeña demora entre el momento en que apartamos la parte del cuerpo afectada y el momento en que se inicia la sensación de dolor; pero explicar la relación causal no nos dice nada sobre la sensación de dolor en sí misma como experiencia subjetiva.

El estado actual del problema mente/cuerpo se basa en dos proposiciones, ambas plausibles, pero aparentemente incompatibles: 1) alguna forma de materialismo o fisicalismo está en lo cierto; 2) el fisicalismo no puede explicar la conciencia fenoménica, la sensación cruda (o *qualia*). Si elegimos la primera, somos materialistas; si, en cambio, preferimos la segunda, somos dualistas. Pero Levine opta por ponerse el mundo por montera y elige las dos: es materialista y «a la vez» cree que los hechos fenoménicos no pueden derivarse de hechos físicos. ¿De verdad puede nadar y guardar la ropa? La

mayoría de los filósofos y neurocientíficos dirían que no; entonces, ¿cómo se las arregla?

Levine desecha la idea —dictada por la intuición— de que los acontecimientos mentales (las tan anunciadas experiencias cualitativas) sean distintos de los acontecimientos físicos; por ejemplo, el retraso en la sensación de dolor es un hecho fenoménico perfectamente explicado por un hecho físico. Pero para Levine esa no es la cuestión. Aunque acepta que la estimulación de las neuronas genera una experiencia fenoménica, y que la conciencia debe de ser realmente un fenómeno físico, por otra parte afirma: «Hay dos características interrelacionadas de la experiencia consciente que se resisten a una reducción explicativa a lo físico: la subjetividad y el carácter cualitativo». Si no podemos salvar esa brecha explicando cómo la estimulación de las neuronas equivale a la experiencia del dolor, Levine sugiere que «debe de ser porque los términos que flanquean la señal de identidad [*sic*] representan en sí mismos cosas distintas». ⁴ Eso suena de manera sorprendente a que Levine ha recurrido a alguna forma de dualismo.

Varios años después, sin embargo, Levine dejó claro que no creía que hubiera una brecha real, un vacío entre las neuronas y la experiencia subjetiva; simplemente estaba señalando el hecho de que no sabemos cómo se podría salvar esa distancia. Obviamente, si se piensa en ello, ese tipo de «brechas» abundan en la historia de la ciencia, pero por regla general se definen como «brechas» o «lagunas» de conocimiento. En última instancia, Levine pensaba que también era ese el caso de la brecha mente/cerebro. En términos filosóficos más sofisticados, creía que se trataba de una cuestión de epistemología *versus* metafísica, y consideraba que se trataba de una laguna en la comprensión actual de cómo se explican esas cosas. Es obvio que, planteado así, está completamente en lo cierto.

Una opinión aún más firme es la del filósofo australiano David Chalmers, que también coincide en que existe una «brecha explicativa». Chalmers está firmemente comprometido con la perspectiva de la segunda de las

proposiciones anteriores: el fisicalismo no puede explicar la conciencia fenoménica, la sensación cruda (o *qualia*), lo que lo convierte en un dualista, aunque él especificaría que es un dualista naturalista. Acepta que los estados mentales están originados por los sistemas físicos del cerebro (esa es la parte naturalista), pero a la vez cree que dichos estados son esencialmente distintos de los sistemas físicos y no reducibles a ellos.⁵ Esta es una postura extraordinaria para un filósofo moderno, pero no para quienes no lo son, ya que actualmente la mayoría de los habitantes del planeta son dualistas.

En cambio Tyndall, en 1879, reformulando ligeramente el discurso inaugural como nuevo presidente de la Asociación Británica que pronunciara en 1874 —y anticipándose a nuestro análisis sobre el origen de la vida, que examinaremos más adelante en este mismo capítulo—, escribía: «Creyendo como lo hago en la continuidad de la naturaleza, no puedo detenerme abruptamente allí donde nuestros microscopios dejan de ser de utilidad. Aquí la visión de la mente complementa con autoridad a la visión del ojo. Por una necesidad intelectual, cruzo el límite de la evidencia experimental y discierno en esa “materia” [...] la promesa y la potencia de toda la vida terrestre».⁶ William James era de la misma opinión al afirmar:

La exigencia de continuidad, en grandes extensiones de la ciencia, ha demostrado poseer una auténtica capacidad profética. En consecuencia, debemos probar sinceramente todas las formas posibles de concebir el surgimiento de la conciencia para que «no» parezca equivalente a la irrupción en el universo de una nueva naturaleza, inexistente hasta entonces.⁷

Asimismo, añadía:

El argumento en el que como evolucionistas debemos mantenernos firmes es que todas las nuevas formas de ser que aparecen en realidad no son más que resultados de la redistribución de los materiales originales e inmutables. Los mismos átomos que, caóticamente dispersos, formaron la nebulosa, ahora, apiñados y temporalmente atrapados en posiciones peculiares, forman nuestro cerebro; y la «evolución» de los cerebros, si se entendiera, sería simplemente el relato de cómo los átomos llegaron a estar tan atrapados y apiñados. En esta historia no se introducen nuevas «naturalezas», ni factores que no estuvieran presentes al principio, en una etapa posterior.⁸

Parece que en las últimas décadas la mayoría de nosotros hemos olvidado que la conciencia humana ha evolucionado gradualmente a partir de nuestros precursores; no surgió plenamente formada en el cerebro del primer *Homo como-se-llame*. James pasaba luego a afirmar: «Si la evolución funciona como debe, la conciencia en una u otra forma tiene que haber estado presente en el mismo origen de las cosas»;⁹ en efecto, tan atrás. Si queremos llegar a entender el abismo entre mente y cerebro, debemos ahondar en otras grandes cuestiones; por ejemplo, cómo es que la vida surge de la materia no viviente.

El viaje de este capítulo nos llevará a descubrir que, para comprender qué diferencia a la materia viviente de la no viviente, es necesario entender la dualidad intrínseca de todas las entidades dotadas de capacidad evolutiva: el hecho de que, en la práctica, toda la materia viviente puede hallarse en dos estados distintos a la vez. Como veremos, la física y la biosemiótica pueden enseñarnos cómo resolver la brecha intrínseca entre los sistemas vivientes y no vivientes sin tener que recurrir a introducir fantasmas en el sistema. Las ideas derivadas de estas disciplinas sugieren cómo concebir el problema de este tipo de brechas en general y cómo determinar la forma de salvar esta en particular, además de ofrecer una hoja de ruta de cómo los neurocientíficos podrían afrontar una brecha mente/cerebro incardinada en una arquitectura estratificada, con protocolos que describen las interfaces entre sus distintas capas. Pero vayamos primero a la física.

Los comienzos de la física y su compromiso con el determinismo

La historia empieza con Isaac Newton y los espectaculares inicios de la física clásica en el siglo XVII. Este es el tipo de física que la mayoría de nosotros nos esforzamos por aprender en la escuela. Resulta que la historia de la manzana es real como la vida misma. El propio Newton se la refirió a su biógrafo,

William Stukeley, recordando un día de 1666 en que, sentado bajo un manzano, se preguntó:

¿Por qué esa manzana siempre tiene que descender perpendicularmente al suelo? [...] ¿Por qué no se desplaza hacia un lado, o hacia arriba, sino siempre hacia el centro de la Tierra? Sin duda la razón es que la Tierra la atrae. Debe de haber un poder de atracción en la materia, y la suma del poder de atracción en la materia de la Tierra debe de estar en el centro de esta, y no en ninguno de sus lados. Es por eso por lo que esta manzana cae perpendicularmente, o hacia el centro. Si la materia atrae así a la materia, debe de ser en proporción a su cantidad. Por lo tanto, la manzana atrae a la Tierra al igual que la Tierra atrae a la manzana.¹⁰

El esposo de la sobrina de Newton, John Conduitt, explicaba que luego Newton pasó a preguntarse si ese poder podría extenderse más allá de la Tierra: «¿Por qué no hasta la Luna?, se dijo a sí mismo, y, de ser así, eso debe de influir en su movimiento y quizá mantenerla en su órbita; después de lo cual se puso a calcular cuál sería el efecto de esa suposición».¹¹ ¡Y vaya si lo calculó! Newton convirtió los resultados de los experimentos sobre el movimiento «terrestre» de Galileo en las ecuaciones algebraicas que hoy conocemos como leyes del movimiento. Galileo había mostrado que los objetos conservan su velocidad y su trayectoria a menos que una fuerza actúe sobre ellos; que los objetos tienen una resistencia natural a los cambios de movimiento, lo que se conoce como inercia; y, finalmente, que la fricción también es una fuerza. Este último hallazgo está presente en la tercera ley: a cada acción le corresponde siempre una reacción igual y de signo contrario. Las reflexiones de Newton sobre la manzana y sus diversos cálculos le llevaron a formular la ley universal de la gravitación y a comprender que las leyes del movimiento «terrestre» a las que él había dado forma algebraica también describían las observaciones realizadas por Johannes Kepler sobre los movimientos de los planetas. No fue un mal día de trabajo.

Así llegó la gran revelación de Newton: acababa de encontrar una serie de relaciones matemáticas fijas y cognoscibles que describían nada menos que el funcionamiento de toda la materia física del universo, desde las bolas de

petanca hasta los planetas. Se trataba de leyes universales e inexorables; independientes de Newton en cuanto observador, de Kepler y de todos los demás. El universo y todos los sistemas que contiene bullen de actividad siguiendo esas leyes —relativas al espacio, el tiempo, la materia y la energía— tanto si hay observadores como si no. Cuando cae un árbol en el bosque sin que nadie lo observe, sigue generando ondas sonoras. Que alguien las oiga o no es otra cuestión, y pronto veremos que esa distinción ilustra el quid de nuestro problema sobre el origen de la vida.

Newton despertó algo más que un mero interés científico. Se dedujo que, si sus leyes eran universales, entonces, teóricamente, si se conocieran las condiciones iniciales, todas las acciones del universo físico serían predecibles. Eso significa que todas las acciones están determinadas, incluso las «nuestras», puesto que solo somos un ente físico más en el universo; basta introducir en la ecuación las condiciones iniciales correctas y obtendremos la respuesta acerca de lo que sucederá después, incluso lo que haremos al salir del trabajo el jueves que viene. Sin embargo, esta línea de pensamiento ignora un hecho crucial. Dentro de un momento descubriremos que la introducción de un determinado valor correspondiente a esas condiciones iniciales es una decisión subjetiva realizada por el experimentador, y esa decisión subjetiva resulta ser un lobo con piel de cordero. Así que las cosas no son tan sencillas.

Las leyes de Newton parecen socavar el libre albedrío y, en consecuencia, la responsabilidad por las propias acciones. Aun así, en un primer momento, el determinismo cautivó la imaginación de los físicos, y pronto muchos otros se vieron también atrapados bajo su influencia. Pero aunque la visión de las cosas de Newton tardó algún tiempo en cuajar, sus leyes parecían describir bien la mayoría de las observaciones del mundo físico, de modo que arraigaron durante los doscientos años siguientes. Sin embargo, no tardó en surgir un nuevo desafío a la física newtoniana, que tenía que ver con un reciente invento: la máquina de vapor. En 1698, un ingeniero militar llamado Thomas Savery patentó la primera versión comercial, destinada a extraer agua

de las minas de carbón inundadas. Aunque luego el diseño de las máquinas fue mejorando, había un problema que parecía no tener solución: la cantidad de trabajo que producían era minúscula comparada con la cantidad de madera que había que quemar para producirlo.

Las primeras máquinas de vapor resultaban ser todas ellas extremadamente ineficientes porque se disipaba o se perdía demasiada cantidad de energía. En el mundo determinista concebido por Newton aquello no tenía demasiado sentido, de modo que los físicos teóricos se verían obligados a afrontar el enigma de la energía aparentemente perdida. Pronto surgiría un nuevo campo de estudio, la termodinámica, y, con él, un cambio de paradigma sobre la naturaleza del mundo, que tiene que ver con la relación del calor y la temperatura con la energía y el trabajo. Al reflexionar sobre todo ello, el ámbito de la física cambiaría para siempre, y el mundo determinista de Newton empezaría a verse de forma algo distinta.

El auge de la mecánica cuántica y la visión estadística de la causalidad

No pasó mucho tiempo antes de que el problema de la máquina de vapor desembocara en las dos primeras leyes de la termodinámica. La primera de ellas afirma que la energía interna de un sistema aislado es constante. En sentido estricto, ello equivale a reiterar la ley de conservación de la energía, que declara que, aunque la energía puede transformarse de una forma a otra, no puede crearse ni destruirse. Esto es absolutamente coherente con el mundo determinista de Newton, pero, por otra parte, no deja de ser una aserción muy limitada, puesto que solo vale para sistemas aislados y cerrados.

Es en la segunda ley donde las cosas se ponen interesantes y estimulantes, y esta tiene que ver con lo que se conoce como entropía. Esta ley revela que el calor, dada su naturaleza, no puede fluir espontáneamente de un punto más frío a otro más caliente. Todavía recuerdo el momento en el que me tocó bregar

con este concepto. Era un frío día de invierno en el Dartmouth College, y yo acababa de recibir a un físico con el que debía mantener una reunión en mi despacho. Él había atravesado a pie todo el campus, una fría caminata a la intemperie que casi le había helado la parka. Comenté alegremente que cada vez que alguien entraba en mi despacho su ropa introducía el frío del exterior, de manera que siempre me cogía frío. Él me miró y me dijo: «Aclaremos la física del asunto. No es que el frío se transfiera a usted, sino que el calor de su cuerpo se transfiere a mí, y, dado que ese calor sale de su cuerpo, usted tiene más frío». Me recordó que la segunda ley de la termodinámica puede ser muy útil incluso para comprender aspectos de la vida cotidiana, y luego añadió que teníamos que contratar a otro físico teórico.

El término *entropía* fue originariamente acuñado en el siglo XIX por el físico alemán Rudolf Clausius para aludir al «calor residual»; mide la cantidad de energía térmica que no se puede utilizar para el trabajo. El frío de la parka del físico había aumentado mi estado de entropía, y, en consecuencia, había menos energía disponible para mantenerme caliente.¹² La segunda ley de la termodinámica complica un poco las cosas, pero, por decirlo de una forma sencilla, la cuestión es que tanto en las parkas como en las máquinas de vapor el intercambio de calor no es reversible. Esta era una idea sorprendente para los físicos de mentalidad newtoniana que creían en un mundo determinista. De repente el tiempo ya no era reversible: la flecha del tiempo fluía en un solo sentido. Esto provocaba que la termodinámica entrase en conflicto con las leyes universales de Newton, que afirmaban que en principio todo era reversible. Esta impactante revelación iría abriéndose paso también en otros ámbitos de pensamiento; incluido, como veremos, el pensamiento relativo a la arquitectura estratificada y a la forma de abordar el problema de la brecha mente/cerebro.

Curiosamente, a mediados del siglo XIX, la teoría atómica —esto es, la teoría que sostiene que la materia está hecha de átomos— había sido aceptada por los químicos, que habían empezado a ponerla en práctica, pero todavía no

había alcanzado un consenso generalizado entre los físicos. Uno de los físicos que se interesaron en ella fue el austríaco Ludwig Boltzmann, quien sería conocido sobre todo por la teoría cinética, que afirma que los gases están formados por un gran número de átomos o moléculas que se mueven constantemente, chocando y rebotando entre sí y contra las paredes del recipiente que los contiene, produciendo así un caótico movimiento aleatorio. Boltzmann convirtió las ideas postuladas por Gassendi en el siglo XVII en ciencia pura y dura, concretamente en el campo científico hoy conocido como mecánica estadística. Partiendo de los tipos de moléculas y sus posiciones, la teoría cinética explicaba las propiedades macroscópicas observables de los gases: presión, temperatura, volumen, viscosidad y conductividad térmica.

En líneas generales, la gran idea de Boltzmann fue definir el desorden de un sistema (su entropía) como el resultado colectivo de todo su movimiento molecular. Asimismo sostenía que, con todos aquellos átomos rebotando de un lado a otro sin orden ni concierto, la segunda ley solo podía ser válida en un sentido «estadístico», no en un sentido literal y determinista; es decir, que no se podía saber si una partícula concreta se transfería o no. Con aquella parka junto a mí, era todo el conjunto de mi sistema el que se estaba desordenando. Como diría Michael Corleone: «No es nada personal; son solo negocios».

Boltzmann causó un gran alboroto entre los físicos que todavía concebían un universo absolutamente determinista y gobernado por las leyes de Newton, los cuales estaban convencidos de que no podía ser un universo estadístico donde lo mejor que cabía hacer era formular meras predicciones. Debido a ello, la teoría de Boltzmann fue objeto de repetidos ataques y, lamentablemente, este llegó a sentirse tan frustrado y deprimido que en 1906 se suicidó mientras estaba de vacaciones con su familia cerca de Trieste, justo antes de que se demostrara de manera inequívoca que su teoría era acertada.

Todavía hoy los físicos se sienten desconcertados por las leyes estadísticas. Por un lado, las leyes de Newton son simétricas con respecto al tiempo y, por ende, reversibles; está claro que en el mundo determinista

definido por Newton lo que va hacia delante también puede ir hacia atrás. Pero, obviamente, no es ese el caso de las leyes estadísticas. ¿Cómo puede ser reversible algo que solo sucede con una determinada probabilidad, no con certeza? No puede; y debido a ello estas dos formas de describir la realidad se hallaban en conflicto: hacía falta un pensamiento nuevo que fuera capaz de gestionar esa dualidad. A pesar de que inicialmente les costó asimilar la física atómica, una vez que la aceptaron los físicos tomaron el testigo y abrieron su mente a lo que este nuevo mundo les mostraba. De buenas a primeras, en 1897, el físico inglés Joseph John Thomson descubrió e identificó la primera partícula subatómica: el electrón. Thomson era tanto un gran físico como un gran profesor: no solo recibió el título de sir y el Premio Nobel por su trabajo, sino que ocho de sus ayudantes de investigación fueron galardonados también con su propio Premio Nobel, al igual que su hijo. Uno de los miembros del grupo era Niels Bohr, que sería quien a la larga postularía el concepto de complementariedad, pero no adelantemos acontecimientos... El caso es que aceptar este nuevo mundo requería un cierto convencimiento.

El físico teórico alemán Max Planck estaba obsesionado con la idea de entropía y con la segunda ley de la termodinámica. En un primer momento optó por creer en su validez «absoluta», y no en la endeble versión «estadística» que defendía Boltzmann. Pero, como paladín de la mecánica newtoniana, Planck era consciente de que la entropía planteaba un problema, dado que dentro del concepto de entropía creciente acechaba la espinosa realidad de la irreversibilidad. Planck aceptaba de hecho la idea de la irreversibilidad, pero anhelaba presentar un derivado riguroso de la ley de la entropía que pudiera justificar su irreversibilidad utilizando leyes clásicas. Como la mayoría de los físicos, ansiaba desesperadamente dar con una descripción física única que lo explicara todo. Y las viejas ideas se resisten a desaparecer.

En 1894 se presentó una oportunidad cuando le encomendaron una tarea especial: optimizar las bombillas, maximizando la luz producida al tiempo que se minimizaba la energía utilizada. Para ello tuvo que abordar el problema de

lo que se conoce como «radiación del cuerpo negro». Podemos hacernos una idea de qué es encendiendo un fuego de campamento. Si acercamos un pincho metálico de kebab al fuego, al cabo de un rato la punta se pondrá al rojo vivo. Si se calienta aún más, el color pasará del rojo al amarillo, luego al blanco y luego al azul. A medida que el interior del pincho se calienta, la superficie comienza a emitir radiación electromagnética en forma de luz, denominada «radiación térmica». Cuanto más caliente está el interior (esto es, cuanto mayor es la energía), más corta es la longitud de onda (y más alta la frecuencia) de la luz emitida; en consecuencia, el color cambia. Los físicos no tardaron en postular un objeto idealizado, un emisor y absorbente «perfecto» que sería de color negro cuando estuviera frío, puesto que toda la luz que cayera sobre él se absorbería por completo.

Ese objeto perfecto se conoce como cuerpo negro, y la radiación electromagnética que emite, como radiación del cuerpo negro. Nadie había podido predecir con exactitud qué cantidad de radiación, y con qué frecuencias, emitiría ese cuerpo negro utilizando las leyes clásicas de la física. Las leyes newtonianas funcionaban bien cuando se emitían las frecuencias de luz más bajas (rojas), pero cuando estas aumentaban las predicciones no podían resultar más erróneas. Después de varios intentos y fracasos utilizando la física clásica pura, Planck acudió a regañadientes a la noción estadística de entropía. Una vez que introdujo la idea de «elementos de energía» y consideró que la energía era «una cantidad discreta compuesta de un número entero de partes iguales finitas»,¹³ pudo formular una ecuación que predecía muy bien la radiación del cuerpo negro.

Ni él ni nadie de su época fue consciente de que su ley de radiación se basaba en una novedad conceptual, un cambio fundamental en la forma de ver el mundo: era la primera incursión en el mundo cuántico. El descubrimiento de Planck también sugería que no había un conjunto de leyes fundamentales ni un modelo del universo. Algunos dirían que aquello representó un paso más hacia la destrucción de la creencia newtoniana de que el mundo era determinista.

Curiosamente, al propio Planck le divirtió la precisión de su ley de radiación, y consideró que postular cuantos de energía era, según sus palabras, «una suposición puramente formal en la que realmente no pensé demasiado». ¹⁴ Lo que Planck había descubierto de forma accidental, la novedad conceptual que no supo entender del todo pero que utilizó como un truco matemático, era que «los objetos microscópicos se comportan de manera distinta de los macroscópicos». ¡Vaya! Planck había quitado involuntariamente el ladrillo que mantenía unidos los cimientos de su sueño más acariciado: que una sola explicación pudiera describirlo todo. El descubrimiento del mundo cuántico no solo cambiaría la comprensión humana del universo, sino que evidenciaría asimismo que existen dos capas de realidad diversas, y que cada una de ellas posee un vocabulario diferente y actúa de una forma distinta. Como en cualquier sistema complejo, cada capa tiene su propio protocolo: en el nivel atómico las cosas funcionaban estadísticamente, mientras que los objetos grandes funcionaban exactamente como decía Newton. Al carecer del concepto de arquitectura estratificada, los epistemólogos se desquiciaron. Los encargados de averiguar «¿cómo es que sabemos?» se sumieron en el caos a base de bien. Lo de una sola explicación para todo no funcionaba. Parecía haber dos clases de explicaciones para el comportamiento de la materia; es decir, una complementariedad.

Los físicos se tropezaron con esa nueva idea cuando comprendieron que la luz podía comportarse como partículas o como ondas. Había una complementariedad en las cosas, una dualidad. Durante décadas lucharon contra aquella idea, pero terminaron aceptándola como verdad. Recientemente, un equipo de investigadores ha captado una imagen increíble de un pequeño grupo de fotones comportándose como ondas y otro grupo comportándose como partículas al mismo tiempo. ¹⁵ Aunque actualmente la idea de complementariedad está bien consolidada en física, en general no se la considera una posible idea básica para concebir la brecha explicativa mente/cerebro. Yo creo que sí debería hacerse, pero primero quiero examinar

cómo la física llegó a aceptar esa realidad aparentemente desconcertante. Tras su aceptación en la física, la idea de complementariedad puede revelarse clave para concebir la biología en general y la brecha mente/cerebro en particular.

La idea de complementariedad

Tras obtener su diploma como profesor de física y matemáticas, en 1901, Albert Einstein, que por entonces tenía veintidós años, se convirtió en ciudadano suizo y se puso a buscar trabajo. Ninguna institución educativa le contrataba, de modo que al final aceptó un puesto en la oficina de patentes de Berna como «experto técnico de tercera clase», labor que compaginaba con clases particulares. En su tiempo libre intercambiaba ideas con un par de amigos en un club de debate que habían formado y al que llamaban Academia Olimpia.

En el transcurso de 1905, que pasaría a conocerse como su *annus mirabilis*, Einstein, ahora con veintiséis años, llevó la física a un nuevo universo al proponer cuatro grandes ideas. Formuló la teoría cuántica de la luz, que estipulaba que la energía de un haz de luz en realidad estaba integrada por pequeños paquetes (más tarde llamados «fotones») y que dicha energía solo podía intercambiarse en minúsculas cantidades discretas. Parecía que, después de todo, el «paquete de energía» no era un mero truco matemático ideado por Planck para producir una buena ecuación. Hasta ese momento se había estado debatiendo si la luz era un movimiento ondulatorio o un conjunto de diminutas partículas. Concebir la luz como onda explicaba toda una serie de observaciones, como la refracción y difracción, la interferencia y la polarización. Pero no explicaba el efecto fotoeléctrico: cuando la luz incide en una superficie metálica, pueden salir proyectados electrones (llamados «fotoelectrones» en este caso) de la superficie del metal.

Inicialmente, los físicos no lo consideraron un gran problema. Partiendo de

la teoría ondulatoria de la luz, supusieron que, cuanto más intensa era esta (es decir, cuanto mayor era la amplitud de onda), mayor era también la energía con la que los electrones salían proyectados del metal. Pero resulta que eso es lo contrario de lo que realmente sucede. La energía de los electrones emitidos es independiente de la intensidad de la luz: tanto las luces brillantes como las mortecinas expulsan electrones de la superficie del metal con la misma energía cuando la frecuencia de las ondas se mantiene constante. La inesperada conclusión resultaba ser que era el aumento de la frecuencia de las ondas lo que incrementaba la energía con la que los electrones eran proyectados desde la superficie. Eso no tiene sentido si la luz es un movimiento ondulatorio. Sería como decir que, si una enorme ola oceánica y una pequeña ondulación alcanzaran una pelota de playa, esta saldría disparada con la misma energía. Einstein comprendió que los efectos observados solo podían explicarse si la luz estaba formada por partículas que interactuaban con los electrones del metal. En su modelo, la luz estaba integrada por cuantos individuales (que más tarde pasarían a denominarse fotones) que interactuaban con los electrones del metal. Cada fotón portaba su propia energía. El aumento de la intensidad de la luz incrementaba la cantidad de fotones por unidad de tiempo, pero la cantidad de energía por fotón seguía siendo la misma. Luego, unos meses después, Einstein acrecentó aún más su año de bonanza determinando que la luz también podía concebirse como onda. La luz existía, pues, en dos realidades simultáneas.

De hecho, Einstein estaba imparable. También presentó evidencias empíricas que validaban la realidad del átomo, zanjando el debate sobre su existencia, y dio el espaldarazo al uso de la física estadística. Por si eso fuera poco, añadió la teoría de la relatividad y formuló la famosa ecuación $E = mc^2$. El mundo de la física tardó un tiempo en asimilar todas aquellas ideas, y en un primer momento Einstein no obtuvo un gran reconocimiento. El resultado inmediato de sus esfuerzos fue únicamente un ascenso en la oficina de patentes a «experto técnico de segunda clase».

Sin embargo, cuando los físicos entendieron la teoría atómica y se pusieron a la altura de los químicos, no tardaron en darse cuenta de que las partículas subatómicas, los átomos y las moléculas —es decir, los submicroscópicos componentes básicos de todo— no seguían las leyes de Newton, sino que las ignoraban abiertamente. Prueba de ello era que, cuando los electrones que orbitaban el núcleo perdían energía, no se estrellaban contra este como habrían predicho las leyes de Newton, sino que se mantenían en órbita. ¿Cómo era posible?

Entre 1925 y 1926, un grupo de físicos —entre ellos Werner Heisenberg, de la Universidad de Gotinga, que realizaba frecuentes viajes de ida y vuelta al instituto de Niels Bohr en Copenhague— desarrollaron la teoría cuántica con el fin de explicar los tres grandes enigmas: el fenómeno de la radiación del cuerpo negro, el efecto fotoeléctrico y la estabilidad de los electrones en órbita. Con ello los físicos, quisieran o no (y muchos, incluidos Planck y Einstein, no querían), se vieron expulsados del mundo determinista de Newton —la «capa» física en la que habitamos y que podemos ver y tocar, el mundo en el que basta una sola explicación para todo— y desplazados a una capa inferior, el mundo oculto, no intuitivo, estadístico e indeterminado de la mecánica cuántica. Fueron expulsados del mundo de las respuestas en blanco y negro al mundo de las respuestas con diversos matices de gris: una capa con un protocolo distinto que existe de manera simultánea.

Consideremos, por ejemplo, la reflexión de la luz. Cuando los fotones inciden en un cristal, un 4 % de ellos se reflejan, mientras que el resto se absorben. ¿Qué determina cuáles de ellos se reflejan? Después de años de investigación empleando múltiples técnicas, la respuesta parece ser esta: el azar. Es el azar el que determina que un determinado fotón sea reflejado o absorbido. Richard Feynman se preguntaba: «¿Nos vemos rebajados, pues, al horror de que la física no se reduzca a predicciones maravillosas sino a probabilidades? Sí, en efecto, esa es hoy la situación [...] Pese al hecho de que los filósofos han dicho: “Es un requisito necesario para la ciencia que la

realización de un experimento exactamente similar produzca resultados exactamente iguales la segunda vez”. De ningún modo. Uno de cada veinticinco sube y, a veces, baja [...] impredecible, absolutamente por azar [...] así son las cosas». ¹⁶ Es el mundo de la incertidumbre, que los físicos de la época despreciaron. Incluso el propio Einstein, que había abierto la puerta a ese mundo incierto, quiso cerrarla de golpe. Albergaba serias dudas sobre lo que implicaba para un universo supuestamente determinista y para la idea de causalidad, lo que le llevó a pronunciar su famosa frase: «Dios no juega a los dados con el universo». No obstante, si pretendían ser buenos científicos, los físicos tenían que descartar sus ideas preconcebidas y estar dispuestos a ir a donde sus descubrimientos los llevaran.

A la hora de explorar el extravagante mundo cuántico, debemos recordar que habitamos el mundo macroscópico de la física newtoniana. El sentido común, esto es, nuestra física popular basada en dicho mundo, no nos ayudará en nada en el mundo cuántico. Este no se parece a nada que hayamos experimentado antes. Debemos dejar nuestra intuición en casa: no nos hará falta; de hecho, solo será una carga. Feynman formuló esta entretenida advertencia cuando preparaba a un grupo de alumnos de física para una clase sobre el comportamiento cuántico:

Su experiencia con lo que han visto hasta ahora es insuficiente. Es incompleta. El comportamiento de las cosas a escala muy pequeña es sencillamente distinto. No se comportan solo como partículas. No se comportan solo como ondas. [...] [Los electrones] se comportan de forma distinta a todo lo que han visto antes. Hay al menos una simplificación: en este aspecto los electrones se comportan exactamente igual que los fotones. Es decir, que ambos son disparatados, pero exactamente de la misma forma. Así pues, se necesita mucha imaginación para apreciar cómo se comportan, puesto que vamos a describir algo que es distinto de cualquier cosa que ustedes puedan conocer. [...] Es abstracto en el sentido de que no es cercano a la experiencia. ¹⁷

Luego continúa diciendo que, si queremos comprender el carácter de la ley física, resulta esencial hablar de este aspecto concreto «porque esto es algo completamente característico de todas las partículas de la naturaleza».

El submicroscópico mundo cuántico está oculto a nuestros ojos. Eso significa que, para aprender algo al respecto, debemos tener algún tipo de interacción que nos permita realizar mediciones, lo que supone involucrar algún instrumento de nuestro mundo macroscópico, el cual a su vez está hecho de átomos, que por su parte pueden perturbar o reaccionar con las partículas que nos proponíamos medir y que iban tranquilamente a su aire. Esa perturbación encaminará la dinámica del sistema en una dirección distinta de la que tenía antes de que se realizara la medición. En resumen: también empezaba a parecer que había un inevitable problema de medición. Fisgonear en el mundo cuántico iba a ser difícil y requeriría un nuevo tipo de pensamiento.

Así que allá vamos: resulta que, tal como descubrió Einstein, la luz se comporta a la vez como onda y como partícula. Unos años después se descubrió que lo mismo vale para la materia: los electrones también tienen propiedades características de las partículas y propiedades características de las ondas. Los físicos pronto aceptaron la idea de que lo que percibimos en nuestro mundo macroscópico como un continuo (en lugar de miles de millones de átomos individuales), pongamos por caso una mesa de comedor, no es más que un proceso simulado basado en promedios en lo que el experto en matemática aplicada, físico y polifacético erudito John von Neumann denominaría más tarde «un mundo que en realidad es discontinuo por su propia naturaleza». Y añadió: «Esta simulación es tal que el hombre generalmente percibe la suma de muchos miles de millones de procesos elementales de manera simultánea, de modo que la ley niveladora de los grandes números oscurece por completo la naturaleza real de los procesos individuales». ¹⁸ Ese efecto «nivelador» de la ley de los grandes números hace que los movimientos de todas esas partículas juntas se anulen entre sí; de ahí que la mesa se quede en su sitio y no vibre desplazándose por el suelo. Aun así, cuando vemos una mesa sólida, en realidad es una mera ilusión, una representación simbólica creada por nuestro cerebro para denotar lo que

realmente está ahí. Es una ilusión muy buena que nos proporciona buena información, lo que nos permite funcionar de manera eficaz en el mundo.

El famoso físico austríaco del «gato en la caja», Erwin Schrödinger, también ansiaba apuntalar el mundo determinista de la causalidad. En esa línea desarrolló la que pasaría a conocerse como ecuación de Schrödinger, una «ley» que describe el comportamiento de una onda en la mecánica cuántica y el modo en que esta cambia dinámicamente a lo largo del tiempo. Aunque dicha «ley» es reversible y determinista, no puede describir el estado general del sistema, ya que no tiene en cuenta el carácter de partícula del electrón, que Schrödinger trataba de eludir. La ley no puede determinar en qué punto exacto de su órbita estará un electrón en un instante en concreto: solo puede hacer una predicción, basada en probabilidades, de cuál sería su posición exacta en ese instante, lo que se denomina su «estado cuántico».

Para conocer la ubicación real del electrón hay que realizar una medición, y ahí es donde empiezan los problemas para los deterministas recalcitrantes. Cuando se realiza una medición, se dice que el estado cuántico «colapsa», lo que significa que todos los demás estados posibles en los que podría hallarse el electrón (conocidos como «superposiciones») se colapsan en uno solo; todas las demás posibilidades desaparecen. Obviamente, la medición es irreversible y constriñe al sistema, lo que provoca su colapso. En los dos años siguientes, los físicos se dieron cuenta de que ni el concepto clásico de partícula ni el de onda podían describir por completo el comportamiento de los objetos de escala cuántica en un determinado instante del tiempo. Como bromeaba Feynman, «No se comportan como una onda o como una partícula; se comportan de forma mecánicamente cuántica».¹⁹

Fue aquí donde vino a echar una mano el físico danés Niels Bohr, experto en electrones y premio Nobel. Tras pasar un par de semanas esquiendo solo en Noruega, reflexionando sobre la doble naturaleza de los electrones y los fotones, regresó con el esbozo del que pasaría a denominarse «principio de complementariedad», que ejemplifica la dualidad onda/partícula. Dicho

principio sostiene que los objetos cuánticos tienen propiedades complementarias que no se pueden medir ni, por ende, conocer en un mismo instante del tiempo. Como describe Jim Baggott en su libro *La historia del cuanto*:

[Bohr] comprendió que las relaciones de incertidumbre posición-momento y energía-tiempo en realidad manifiestan la complementariedad entre los conceptos clásicos de onda y de partícula. El comportamiento como onda y el comportamiento como partícula son inherentes a todos los sistemas cuánticos expuestos a experimentos, y al elegir un experimento —al elegir el espejo de las ondas o el espejo de las partículas—, introducimos una inevitable incertidumbre en las propiedades que hay que medir. No se trata de una incertidumbre introducida por la «tosquedad» de nuestras mediciones, como argumentara Heisenberg, sino que surge porque nuestra elección de un determinado aparato obliga al sistema cuántico a revelar un tipo de comportamiento por encima de otro. ²⁰

Una vez más: en cualquier instante en concreto se puede medir y conocer la posición o el momento ²¹ de un electrón, pero no ambos, al igual que ocurre con sus propiedades como onda o como partícula. Cuando realizamos una medición instantánea de su posición en un único punto del tiempo, el electrón se encuentra de hecho en una única ubicación y no se mueve; en consecuencia, su doble naturaleza —que le dota también de un determinado momento, además de una posición— se ve comprometida. En ese instante no es posible medir el momento: solo puede sugerirse la otra medida con un cierto grado de probabilidad, no con certeza. La complementariedad emerge en un sistema cuando se intenta medir una de las propiedades emparejadas. Un mismo sistema tiene dos modos simultáneos de descripción, ninguno de los cuales es reducible al otro.

Bohr trabajó durante seis meses en esta teoría, que expuso por primera vez en una conferencia pronunciada en 1927 en la ciudad italiana de Como, donde la presentó a un ilustre grupo de físicos en el marco de un congreso conmemorativo del centenario de la muerte de Alessandro Volta. Einstein, que no estaba allí, no supo de su existencia hasta el mes siguiente, cuando Bohr volvió a presentarla, esta vez en Bruselas. A Einstein no le gustó la idea de la

doble descripción y la incertidumbre, y Bohr y él iniciaron un largo intercambio epistolar que duraría años. En su intento de derrotar a la teoría cuántica, Einstein le presentaba a Bohr un determinado escenario que parecía refutarla, solo para encontrarse con que este le presentaba a su vez un argumento coherente con la teoría cuántica que le permitía salir airoso. Desde entonces se han formulado muchas propuestas y se han realizado experimentos para reforzar la posición del bando de Einstein en este debate;²² todos sin éxito. Por impopular que resulte entre quienes se decantan por un sesgo determinista, hay una versión de la complementariedad de Bohr que permanece invicta.

Existen dos conceptos esenciales en el debate entre ambos físicos: qué significa la objetividad y de qué trata la física. Robert Rosen ha explicado qué es exactamente lo que está en juego:

La física se esfuerza cuando menos en limitarse a «objetividades». En consecuencia, presupone una rígida separación entre lo que es objetivo y, por ende, cae directamente dentro de sus dominios, y lo que no lo es. Su opinión sobre lo que quiera que quede fuera de tales dominios está dividida. Algunos creen que lo que queda fuera lo está debido a cuestiones técnicas de formulación eliminables y provisionales; es decir, que lo que queda fuera puede «reducirse» a lo que ya está dentro. Otros creen que esa separación es absoluta e irrevocable.²³

Bohr, que pertenecía al último de esos dos grupos, argumentaba que el hecho de que veamos la luz como partícula o como onda no es inherente a la propia luz, sino que depende del modo en que se la mida y se la observe. Tanto la luz como el aparato de medición forman parte del sistema. En opinión de Bohr, el mundo clásico es demasiado pequeño para describir toda la realidad material; él consideraba que el universo y todo lo que contiene son mucho más complejos y requieren algo más que una única capa con un protocolo formado por las leyes de la física clásica. Rosen afirma que Bohr cambió el propio concepto de «objetividad», que pasó de referirse únicamente a lo que es inherente de forma exclusiva a un sistema material a definir lo que es inherente a un par sistema-observador. Einstein no pudo digerir esa idea y

optó por unir su suerte a la de la física clásica, que ignora el procedimiento de medición y considera su resultado inherente a la luz. Para él, algo es objetivo solo si es independiente de cómo se mida u observe. Concluye Rosen: «Einstein creía que existía tal conocimiento, inmanente solo en la cosa, e independiente de cómo se obtuviera. Bohr consideraba esa visión “clásica”, incompatible con las visiones cuánticas de la realidad, que siempre requerían la especificación de un contexto y siempre contenían información indivisible [*sic*] sobre ese contexto». ²⁴

La importancia del principio de complementariedad de Bohr no reside solo en su interés como objeto de su disputa científica con Einstein. A continuación veremos que es fundamental para entender la brecha mente/cerebro.

8

De lo no viviente a lo viviente y de las neuronas a la mente

En el principio del cielo y la tierra no había símbolos. Los símbolos salieron del útero de la materia.

LAO TSE

Se habrá alcanzado una etapa importante del pensamiento humano cuando lo fisiológico y lo psicológico, lo objetivo y lo subjetivo, estén realmente unidos.

IVÁN PÁVLOV

En su intento de comprender la materia de forma más precisa, los físicos se tropezaron con la idea de complementariedad, el concepto de que toda la materia puede existir en dos estados distintos a la vez. Aceptar esta dualidad implicaba algo más que el mero hecho de ampliar los límites de la física: comprender el mundo natural exigía ahora un nuevo pensamiento que fuera más allá de lo que podíamos concebir a partir de nuestra propia experiencia de los fenómenos naturales. De manera similar, quienes estudian hoy la dualidad mente/cerebro necesitan también un nuevo pensamiento y un esfuerzo de imaginación. Hace falta alguien capaz de pensar más allá del marco mental de los últimos dos mil quinientos años, lleno de intuiciones y saber convencional. Hace falta alguien que haya bregado con los altibajos de la física moderna y reconozca la importancia de la complementariedad. Hace falta alguien que

crea que los filósofos han desperdiciado los últimos milenios buscando respuestas en el lugar equivocado: un órgano extremadamente evolucionado como el cerebro humano. Hace falta alguien como Howard Pattee, un físico educado en Stanford que se metió de lleno en la biología teórica durante su impresionante carrera en la Universidad de Binghamton (una de las cuatro que integran el sistema de la Universidad Estatal de Nueva York, o SUNY, por sus siglas en inglés). Pattee, un atento observador del pensamiento humano, considera que los filósofos han abordado la división mente/cerebro desde el extremo equivocado de la evolución.¹ A lo largo de su trayectoria profesional llegó a una sorprendente conclusión: la dualidad es una propiedad intrínseca y necesaria de cualquier entidad capaz de evolucionar.

A Pattee no le preocupa en concreto la brecha entre el cerebro material y la mente inmaterial. Él profundiza mucho más. La brecha originaria, la verdadera fuente del problema, ya estaba ahí mucho antes de que existiera el cerebro: la madre de todas las brechas es la que existe entre la materia viviente y la no viviente. El problema fundamental surgió en el origen de la vida en la Tierra. Así pues, lejos de centrarnos exclusivamente en la división entre el cerebro físico y la mente etérea, lo que necesitamos es más bien entender la diferencia entre los conglomerados de materia que producen un objeto sin vida y los conglomerados de materia que producen algo que está lleno de ella. La diferencia entre lo viviente y lo no viviente está en la raíz de la división entre el cerebro y la mente, y ofrece un marco adecuado para abordar el problema de la dualidad mente/cerebro.

Asimilar la idea de Pattee requiere cierto tiempo e implica asimismo tener en cuenta una importante advertencia: si queremos llegar a comprender el concepto de conciencia, algo plenamente formado en los sistemas vivientes evolucionados, primero necesitamos entender qué es lo que hace de entrada que un sistema viviente esté vivo y sea capaz de evolucionar. ¿Qué fue lo que dividió las cosas en dos dominios distintos, uno viviente y el otro no? La mayoría de nosotros hemos dedicado tan solo unos segundos a pensar en el

problema de «cómo la vida surgió de la materia», pero pronto lo hemos dejado correr por considerarlo demasiado complejo, para pasar a dedicarnos a explorar y medir lo que tenemos ante nuestros ojos. No así Pattee. Él se sintió cautivado por las preguntas relativas al origen de la vida ya de adolescente, y empezó a chapotear en aguas científicas a finales de la década de 1930, mientras estaba en un internado. Su profesor de ciencias y director de la institución, el doctor Paul Luther Karl Gross, le recomendó un libro para que lo leyera durante el verano. No era precisamente el tipo de lectura ligera característica de las vacaciones estivales: se trataba de *La gramática de la ciencia*, del brillante matemático y estadístico británico Karl Pearson, una obra publicada por primera vez en 1892.

En aquel momento, Pattee se preguntó por qué le recomendaban un libro de ciencia aparentemente obsoleto escrito antes de los días de la teoría cuántica. Sin embargo, en el capítulo titulado «La relación de la biología con la física», Pattee encontró una pregunta que se convertiría en el *leitmotiv* de su pensamiento durante décadas: «¿Cómo [...] diferenciar lo viviente de lo inerte si resulta que podemos describir conceptualmente ambas cosas mediante el movimiento de corpúsculos inorgánicos?». ² Pattee supo ver la lógica de la pregunta, pero también comprendió que recurrir a las mismas leyes para describir tanto la materia animada como la inanimada no podía dar lugar a una explicación lo suficientemente buena. De hecho, no conducía a ninguna explicación en absoluto. Tenía que haber algo más.

El quebradero de cabeza de la mecánica cuántica

Pattee fue muy afortunado por tener a un profesor y director tan excepcional. El doctor Gross estimulaba el pensamiento de sus alumnos llevándolos a eventos científicos de vanguardia, incluyendo las conferencias vespertinas que daba el premio Nobel Linus Pauling en el Instituto Tecnológico de California. Una tarde, Pattee escuchó a Pauling describir la famosa paradoja del gato de

Schrödinger, que implicaba la idea de que un gato podía estar vivo y muerto a la vez. La paradoja era esta: un gato está encerrado en una cámara de acero junto con una pequeña cantidad de material radiactivo y un contador Geiger para medir la radiación que emite. Dada la velocidad de desintegración radiactiva del material, existe un 50 % de probabilidades de que en el plazo de una hora ninguno de los átomos se desintegre. Sin embargo, obviamente, también existe otro 50 % de que algún átomo sí lo haga, y en ese caso el gas del tubo del contador Geiger se descargará. En este experimento mental, la descarga del contador Geiger libera un martillo que rompe un frasquito de ácido cianhídrico, que mata al gato. Así pues, esta elaborada configuración produce un escenario en el que hay un 50 % de probabilidades de que al cabo de una hora el gato esté vivo y otro 50 % de que esté muerto. Extraño, pero bastante bueno. Sin embargo, en mecánica cuántica este fenómeno no se expresaría mediante la probabilidad de dos resultados distintos, sino mediante la denominada «función de onda», una descripción de todo el estado cuántico del sistema. Y la función de onda del pobre gato de Schrödinger se traduciría en que habría un gato vivo y un gato muerto a partes iguales. Pattee se quedó perplejo. ¿Cómo podía la mecánica cuántica, una teoría que explicaba toda la química y la mayor parte de la física, producir un absoluto disparate como el problema del gato de Schrödinger? Aquella pregunta marcaría para él el comienzo de un viaje que duraría toda su vida en busca de la resolución de aquel dilema.

El rompecabezas que dejó perplejo al joven Pattee se conoce como el «problema de la medición». Ya hemos comentado en el capítulo anterior que un sistema cuántico tiene propiedades complementarias emparejadas que no se pueden medir de forma simultánea. Hay tres razones por las que en el nivel cuántico las mediciones plantean desafíos adicionales. Para empezar, la medición requiere un observador, un sujeto o agente que sea independiente del objeto que se mide. En segundo término, el proceso de medición (que es irreversible) no se rige por las leyes clásicas de la física. En tercer lugar, la

medición tiene aspectos arbitrarios: el observador elige cuándo, dónde y qué medir, además de los símbolos (en sí mismos arbitrarios) utilizados para expresar la medida obtenida. De hecho, la medición es un proceso selectivo en el que se ignoran la mayoría de los aspectos de lo que se mide. Pongamos por caso que yo quisiera describirle a usted, lector. ¿Qué medida debería elegir para reflejarlo? Supongamos que decido medir su masa: haré una medición de su peso y la utilizaré para describirle a lo largo de toda su vida. ¿Cuándo debería realizarla: cuándo era un bebé; un adulto de veinte años, de treinta y cinco, de sesenta...? ¿El día antes o después de la comida de Navidad? ¿Cuál es más representativa? ¿El peso por sí solo es un buen indicador sobre usted? ¿Y si midiera tanto el peso como la estatura?... Puede que la medición en sí sea precisa y objetiva, pero el «proceso» de medición es subjetivo.

Si el proceso de medición es arbitrario, eso significa que no puede describirse mediante leyes objetivas, ya sean cuánticas o clásicas, lo cual representa un problema para toda la física, no solo para su versión cuántica. Para realizar predicciones sobre el estado futuro de un sistema, el físico debe conocer las condiciones iniciales de este. ¿Cómo? Obviamente, midiendo dichas condiciones. Pero esa medición es arbitraria, y, al hacerla, el físico interfiere con esas mismas condiciones iniciales. Esta subjetividad de la medición inicial a menudo es ignorada por los deterministas cuando suponen que el mundo es completamente predecible. Pero no hay escapatoria. No importa cuánto te esfuerces en ser un observador objetivo: por el mero hecho de medir estás introduciendo subjetividad en el sistema. El problema de la medición supone un duro golpe para la física, pero podría ser justo lo que la neurociencia necesita.

El Schnitt y el origen de la vida

Para referirse a la ineludible separación entre el sujeto (el medidor) y el objeto (lo medido), los físicos suelen utilizar el término *Schnitt* (una sonora

palabra alemana que significa «corte»). Pattee denomina corte epistémico a «esta inevitable separación conceptual entre el conocedor y lo conocido, o entre el registro simbólico de un acontecimiento y el acontecimiento en sí». ³ Hay todo un mundo de acciones en el lado del observador en relación con el registro de un acontecimiento por parte de este, y hay asimismo todo un mundo de acciones, independiente del anterior, en el lado del acontecimiento en sí. Esto parece confuso, pero recuerde la brecha explicativa que existe entre nuestra experiencia subjetiva de un acontecimiento («me he divertido mucho haciendo *bodysurf*») y el acontecimiento en sí (una persona ha ido a nadar al mar). Otra alternativa es pensar en la brecha explicativa que existe entre esa misma experiencia subjetiva («esto es divertido») y lo que sucede dentro del cerebro (algunas neuronas se han activado mientras una persona nadaba en el mar). Todas estas son solo distintas versiones de la complementariedad sujeto/objeto que se observa en física. Aquí está la parte realmente peliaguda: ¿quién mide los acontecimientos? Para examinar la diferencia entre la experiencia subjetiva de una persona y la realidad objetiva, ¿necesitamos a un científico? ¿Y quién mide al científico?

Pattee señala que ni la teoría clásica ni la cuántica definen formalmente al sujeto, es decir, al agente u observador que determina lo que se mide. La física, pues, no dice «dónde» hacer el corte epistémico. ⁴ Sin embargo, la medición cuántica no requiere la presencia de un fisico-observador: Pattee argumenta que hay otras cosas que pueden realizar mediciones cuánticas. Así, por ejemplo, las enzimas (como las ADN polimerasas) pueden actuar como agentes de medición, realizando mediciones cuánticas durante el proceso de replicación celular. ⁵ No hace falta un observador humano.

El sinsentido del gato de Schrödinger, del que ni siquiera el adolescente Pattee quiso hacerse partícipe, no surgía del gato en sí o del contador Geiger, sino del humano que realizaba el extraño experimento. En el experimento mental de Schrödinger, el gato se describe como una función de onda, vivo y muerto a la vez. Esta situación persiste hasta que abrimos la caja y hacemos

una medición; es decir, observamos si el gato está vivo o muerto, pero ya no ambas cosas. El resultado de la intervención en forma de medición (el humano que abre la caja) parece ser instantáneo e irreversible, y la representación física de dicho resultado (gato vivo o gato muerto) es arbitraria. Pero ¿cómo puede ser eso cuando, al mismo tiempo, se supone que todos los acontecimientos microscópicos obedecen a leyes dinámicas cuánticas reversibles (como, por ejemplo, la ecuación de Schrödinger)? Pattee señala que era este modelo inadecuado de medición el que impedía que se conociera el estado del gato antes de ser observado. Y afirma: «Era la creencia de que la conciencia humana colapsaba en última instancia la función de onda la que generaba el problema del gato de Schrödinger». ⁶ De hecho, Schrödinger concibió su experimento del gato precisamente para poner de manifiesto que aquella idea era ridícula: esperaba ilustrar que la superposición cuántica no podía funcionar con objetos grandes, como gatos (o perros, o, para el caso, usted mismo).

Para Schrödinger, la ironía iba sobre nosotros: trataba de hacernos ver que nuestra comprensión es limitada. Pattee supo entenderlo (en el instituto) y se puso a trabajar para abordar el problema. ¿Dónde deberíamos situar el corte, la brecha, el *Schnitt*? Con su apasionado interés por el origen de la vida, se dio cuenta de que la conciencia humana era una capa demasiado elevada en la arquitectura de todos los organismos vivientes para situar en ella el corte epistémico entre el observador y lo observado, entre la experiencia subjetiva y el acontecimiento en sí. Entre las partículas subatómicas y el cerebro humano existen infinidad de capas; de hecho, hay ya un montón de ellas entre las partículas subatómicas y cualquier cerebro en general (de gato o de ratón, de mosca o de gusano). Situar tan arriba el crucial corte epistémico conducía al sinsentido de la existencia del gato de Schrödinger como un sistema cuántico. Pattee no se andaba con rodeos: «He adoptado el punto de vista de que la cuestión de qué constituye una observación en mecánica cuántica debe surgir mucho antes de que lleguemos a la complejidad del cerebro. De hecho,

yo propongo [...] que la brecha entre el comportamiento cuántico y el clásico es inherente a la distinción entre materia inanimada y materia viva». ⁷

He aquí, pues, la idea: Pattee postula que la brecha fue el resultado de un proceso equivalente a la medición cuántica que se inició con la autorreplicación en el origen de la vida y con la célula como el más simple de los agentes. ⁸ El corte epistémico, el corte sujeto/objeto, el corte mente/materia... todos tienen sus raíces en ese corte original producido en el origen de la vida. La brecha entre el sentimiento subjetivo y los estímulos neuronales objetivos no se produjo con la aparición del cerebro: ya estaba allí cuando la primera célula empezó a vivir. Hay dos modos de comportamiento complementarios, dos niveles de descripción, inherentes a la propia vida; estaban presentes en su origen, la evolución los ha preservado y siguen siendo necesarios para diferenciar la experiencia subjetiva del acontecimiento en sí. Es una idea alucinante.

Una vida en símbolos: Von Neumann muestra el camino

La materia viva parece jugar a un juego completamente distinto del de la materia no viviente a pesar de que ambas están hechas del mismo material. ¿Por qué la materia viva es distinta de la no viviente? ¿Puede ser simplemente porque esté haciendo trampas, violando de alguna manera las leyes físicas que, según hemos llegado a entender, gobiernan la materia no viviente? Pattee sostiene que la materia viva se diferencia de la no viviente por su capacidad de reproducirse y evolucionar a lo largo del tiempo. Si es así, ¿qué hace falta para reproducirse y evolucionar?

John von Neumann fue un genio matemático y un electrizante *bon vivant* de origen húngaro cuyas aportaciones intelectuales fueron tan inmensas como su apetito por la vida. Nacido en la aristocracia judía de Budapest, moriría tras recibir los últimos sacramentos de un sacerdote católico, afirmando en broma que había aceptado la apuesta de Pascal. ⁹ En los años transcurridos entre uno

y otro hecho, entre otras cosas se incorporó al Instituto de Estudios Avanzados de Princeton, donde se dice que volvía loco a Einstein poniendo marchas alemanas en su gramófono a todo volumen.

La atmósfera intelectual de la época era vibrante. En 1943, Schrödinger había pronunciado en Dublín su histórica conferencia «¿Qué es la vida?», donde sugirió que, de algún modo, había un «guion codificado» incorporado a los mecanismos moleculares de la célula. A finales de la década de 1940, Von Neumann también se había asignado a sí mismo la tarea de reflexionar sobre la vida como experimento mental. ¿Qué es la vida? O bien ¿qué hacen los seres vivos? Una respuesta es que se reproducen: la vida engendra más vida. Pero la lógica le decía que «lo que ocurre es en realidad de un grado superior a la autorreproducción, ya que los organismos parecen haberse ido haciendo más elaborados con el transcurso del tiempo».¹⁰

La vida «no solo» engendraba más vida, sino que podía incrementar su complejidad: podía evolucionar. Von Neumann se interesó cada vez más en la cuestión de qué requeriría —según la lógica— una máquina autónoma, autorreplicante y capaz de evolucionar (un «autómata») cuando se la colocaba en un entorno con el que pudiera interactuar. Su cadena lógica le llevó a la conclusión de que el autómata necesitaba una descripción de cómo copiarse a sí mismo y una descripción de cómo copiar esa descripción para poder transferirla al siguiente autómata recién creado. El autómata original también necesitaba un mecanismo que le permitiera realizar físicamente el trabajo de construcción y copia. Necesitaba información y construcción. Sin embargo, eso solo cubría el aspecto de la replicación. Von Neumann razonó que había que añadir algo más para que el autómata pudiera evolucionar, incrementar su complejidad; y concluyó que necesitaba una autodescripción simbólica, un genotipo, esto es, una estructura física independiente de la estructura que describía o fenotipo. Vincular la descripción simbólica a aquello a lo que hacía referencia requeriría un código, y con ello sus autómatas podrían evolucionar. Enseguida veremos por qué.

Resulta que Von Neumann dio en la diana, ya que predijo correctamente cómo se replican las células antes de que lo descubrieran Watson y Crick. Desde el principio, en el origen de la vida, a nivel monomolecular, cuando el ADN no era más que un proyecto de la Madre Naturaleza, la autorreplicación evolutiva dependía de dos cosas: 1) la escritura y lectura de registros hereditarios, que adoptaban algún tipo de forma simbólica, y 2) una nítida distinción entre los procesos de descripción y construcción. Tras este pequeño experimento mental, Von Neumann pasó a dedicarse a otros enigmas y tareas. Dejó pues su trabajo a medio hacer, puesto que no abordó la cuestión de cuáles eran los requisitos físicos necesarios para implementar su lógica. Frotándose las manos, Pattee recogió el testigo.

La física de los símbolos: Pattee sigue adelante

Tendemos a pensar que los símbolos son abstractos, no algo sometido a la física. Sin embargo, como científicos, somos seres físicos que buscan evidencias físicas que obedezcan a leyes y reglas físicas. Debe de haber una manifestación física de los símbolos de Von Neumann. Pero lo que Pattee denomina la «física de los símbolos» genera ciertos problemas. El primero radica en la escritura y lectura de los registros hereditarios, la descripción informativa. Una descripción implica un proceso de registro y, como hemos aprendido en el apartado anterior, un registro es una medición irreversible que requiere algún medidor. Pattee comprendió que la descripción informativa en el origen de la vida se topa de frente con el problema de medición de la mecánica cuántica. Las mediciones son subjetivas, lo que significa que no pueden describirse mediante leyes objetivas, ya sean cuánticas o clásicas. Cualquier ser viviente que «registre» información está introduciendo una forma de subjetividad en el sistema.

El segundo problema es la relación entre el genotipo y el fenotipo. Por ejemplo, si consideramos el ADN, el genotipo es la secuencia que contiene las

instrucciones para el organismo vivo, mientras que el fenotipo lo forman las características observables de un organismo, como su anatomía, bioquímica, fisiología y comportamiento. El genotipo interactúa con el entorno para producir el fenotipo. Para trasladar esto a un ejemplo de la vida cotidiana, considere que el proyecto de una casa es su genotipo, y la casa real su fenotipo. El proceso de construcción fenotípica es la construcción de la casa utilizando el proyecto como información acerca de qué hacer y cómo hacerlo. El fenotipo está relacionado con el genotipo que lo describe, pero hay una enorme diferencia física entre el fenotipo y el genotipo, e incluso entre este y el proceso de construcción fenotípica. Para empezar, el genotipo no es dinámico: es una secuencia de símbolos inactiva y unidimensional (los símbolos del ADN son nucleótidos) que no tiene restricciones de energía o de tiempo. Al igual que un proyecto, puede permanecer inmóvil durante años, como probablemente ya habrá aprendido el lector viendo la serie *CSI*. El genotipo dicta qué debe construirse (pongamos por caso, un perrito muy mono), pero el ADN en sí no tiene el aspecto ni actúa como un perrito mono. Por su parte, el fenotipo (el perrito mono en sí) es dinámico y consume energía, especialmente si es un border collie.

El proceso de construcción fenotípica también está relacionado con el genotipo. Al igual que un proyecto impone límites al constructor y le impide agregar torres a una casa, el genotipo limita cuántas colas tendrá el perrito mono. ¿Cómo se relacionan uno y otro? ¿Cuál es la relación entre un proyecto, una casa y todo el trabajo intermedio de verter hormigón, clavar clavos, etc.? Los registros hereditarios contienen no solo la información que especifica «qué» hay que construir, sino también la que especifica «cómo» construirlo. De algún modo, esa información sobre «qué» y «cómo» se ha «registrado» en algún tipo de forma simbólica. Hay una distancia, una brecha, entre el símbolo subjetivamente registrado (genotipo), por una parte, y el proceso de construcción fenotípica y el fenotipo, por otra. Los símbolos deben traducirse a su significado para que se inicie la construcción. Si pensamos en una

arquitectura estratificada, estaríamos hablando del protocolo de comunicación entre dos capas. Pattee postula que el corte epistémico surgió a partir de la interfaz de control entre esas dos capas. En el caso del proyecto de la casa, el puente es el contratista que traduce dicho proyecto para los trabajadores que la construyen; en el del ADN, el puente entre el genotipo y el fenotipo es el código genético.

Pattee amplió la lógica de Von Neumann argumentando que los propios símbolos que componen las instrucciones (el registro hereditario) deben tener una estructura material, y que durante el proceso de construcción fenotípica (la construcción del nuevo autómatas) esa estructura material constriñe el proceso de una forma que sigue las leyes de Newton. No hay ningún truco de magia. Los símbolos son estructuras físicas reales, una cadena de nucleótidos que obedecen las leyes de la física clásicas.

Aquí está la pega: todo símbolo, ya sea una secuencia de nucleótidos en el ADN, una secuencia de código Morse o una secuencia de simulaciones mentales, es de naturaleza arbitraria. Es fácil captar la arbitrariedad de los símbolos cuando se observa un mundo cambiante como el de las jergas idiomáticas. Por ejemplo, *parné*, *plata* o *pasta* han sido todos ellos símbolos populares, aunque arbitrarios, del dinero. Y además cada idioma tiene su propio conjunto de símbolos, como advirtió en cierta ocasión el cómico Steve Martin a cualquiera que tuviera la intención de viajar a París: «*Chapeau* significa “sombrero”; *oeuf* significa “huevo”. ¡Parece que esos franceses tienen una palabra distinta para todo!». ¹¹ El problema es que Newton no tenía nada de arbitrario. Si los símbolos estuvieran regidos por las inflexibles leyes newtonianas, cada persona, todo el mundo, usaría exactamente la misma palabra para representar el concepto de dinero, en todas las ocasiones y por los siglos de los siglos. Por desgracia para Newton, hay numerosos símbolos posibles que pueden elegirse para transmitir información. Cada uno de ellos tendrá propiedades distintas, diferentes ventajas y desventajas, pero, dado que

los símbolos son diversos de la cosa en sí, no existe una correspondencia unívoca.

Se podría objetar que el ADN es arbitrario de la misma forma que lo es el lenguaje, argumentando que existen restricciones fisicoquímicas. Pero la elección de un símbolo no se rige por «leyes» físicas, sino por una «regla»: elegiremos el símbolo que contenga la información más útil y fiable (estable) para el sistema. Enseguida veremos que los propios componentes del ADN fueron seleccionados de entre un abanico de competidores debido a que funcionan mejor a la hora de constreñir la función del sistema del que forman parte; y, si este es estable, eso significa que puede ser transmisible. Los actuales componentes del ADN son lo que Pattee denomina «accidentes congelados». El símbolo actual incorpora la historia de sus versiones exitosas en múltiples escalas de tiempo (es decir: es independiente del tiempo), no la de su acción actual. Volvamos al ejemplo del dinero: si solo un par de personas en el mundo llamaran al dinero «cachumbo», este no sería un símbolo fiable, y no sería seleccionado ni transmitido.

Esto resulta un poco confuso porque en nuestro mundo social solemos mezclar los conceptos de regla y de ley; por ejemplo, cuando se promulga una «ley» estableciendo las «reglas» de conducción de un determinado país. Pattee explica que en la naturaleza existe una distinción básica y extremadamente importante entre leyes y reglas.¹² Las leyes son inexorables, lo que significa que son inmutables, ineludibles e inevitables. No podemos alterar ni evadir las leyes de la naturaleza. Estas dictan que, por ejemplo, un automóvil se mantenga en movimiento hasta que una fuerza igual y de sentido contrario lo detenga o se quede sin energía. Eso no es algo que podamos cambiar. Asimismo, las leyes son incorpóreas, lo que significa que su ejecución no requiere de ninguna encarnación o estructura: no hay un policía de la física que obligue al automóvil a detenerse cuando se queda sin energía. Las leyes también son universales: rigen en todo momento y lugar. Las leyes del movimiento se aplican tanto si estás en Escocia como en España.

Por otro lado, las reglas son arbitrarias y pueden cambiarse. En las islas Británicas, la regla de conducción dicta que debe hacerse por el lado izquierdo de la carretera, mientras que en la Europa continental la regla es conducir por el lado derecho. La ejecución de las reglas sí depende de algún tipo de estructura o de mecanismo restrictivo. En este caso, la estructura es una fuerza policial que multa a quienes incumplen las reglas conduciendo por el lado incorrecto. Las reglas son de carácter local, lo que significa que solo pueden existir donde y cuando haya estructuras físicas que impongan su cumplimiento. Si vives en el interior despoblado de Australia, tú eres quien manda: puedes conducir por el lado que quieras. No hay ninguna estructura que pueda imponerte una restricción. Además de locales, las reglas son modificables y frágiles. Un símbolo, regido por reglas, es seleccionado de entre un abanico de competidores debido a que funciona mejor a la hora de constreñir la función del sistema del que forma parte, lo que se traduce en la producción de un fenotipo con más probabilidades de éxito. La selección es flexible; las leyes de Newton, no. En su papel informativo, los símbolos no dependen de las leyes físicas que gobiernan la energía, el tiempo y los ritmos de cambio. No siguen ninguna de las leyes de Newton: son solo anárquicos seguidores de reglas. Lo que esto nos indica es que los símbolos no están atrapados en sus significados.

Los símbolos llevan una doble vida, puesto que tienen dos modos de descripción diversos y complementarios en función del trabajo que realizan. En una de esas «vidas» los símbolos están hechos de material físico (el ADN está compuesto de moléculas de hidrógeno, oxígeno, carbono, nitrógeno y fosfato) que sigue las leyes de Newton y constriñe el proceso de construcción mediante su estructura física. Sin embargo, en su otra «vida», es decir, en cuanto depósitos de información, los símbolos escapan a esas leyes. Esta doble vida de los símbolos ha sido en gran medida ignorada. Las personas interesadas en el procesamiento de la información ignoran el lado material objetivo, la manifestación física del símbolo. Por su parte, los biólogos

moleculares y los deterministas, interesados solo en el aspecto material, ignoran el lado simbólico subjetivo. Al limitarse únicamente a uno de los dos aspectos, ninguno de ellos estudia su naturaleza íntegra y complementaria, lo cual no es solo una vergüenza, sino también una parodia científica, puesto que —como hemos visto—, para que exista una forma de vida autorreproducible y capaz de evolucionar, los símbolos físicos deben realizar ambos papeles. Pattee sostiene que ninguno de los dos basta por sí solo: evitar cualquiera de los dos extremos del eslabón hace que este se pierda por completo. Y afirma con valentía: «Es precisamente esta articulación natural símbolo-materia la que diferencia la vida de los sistemas físicos no vivientes». ¹³

El código genético es un código real

Para comprender mejor esta articulación símbolo-materia y cuáles son sus implicaciones para nuestro propósito, examinaremos de cerca el ADN, que es el que mejor ejemplifica la estructura símbolo-materia en un sistema viviente. Primero, no obstante, necesitamos aprender algunos conceptos básicos de biosemiótica para entender cómo funcionan los símbolos en los sistemas vivientes. Nuestro guía será Marcello Barbieri, un biólogo teórico de la Universidad de Ferrara.

La semiótica es el estudio de los signos (o, lo que es lo mismo, de los símbolos) y sus significados. Una idea básica en este campo es que, por definición, un signo siempre está vinculado a un significado. Como ya hemos podido deducir por las palabras de Steve Martin y sus problemas en París, no existe una relación determinista entre un signo y su significado. Un huevo es un huevo, tanto si vives en Estados Unidos como en Francia, pero podemos denominarlo con diferentes nombres. El objeto es distinto de su representación simbólica (el sonido *oeuf* en francés, *egg* en inglés, *huevo* en español, etc.), así como de nuestra comprensión del símbolo. Barbieri señala que la relación entre un signo y su significado viene establecida por un código, un conjunto de

reglas convencionales que establece la correspondencia entre los signos y sus significados. Ese código ha sido generado por un agente, un «codificador». Un sistema semiótico se origina cuando el codificador crea el código. Así pues — señala Barbieri—, «un sistema semiótico es una tríada integrada por signos, significados y código, todo ello producido por el mismo agente, es decir, por el mismo codificador». ¹⁴

La biosemiótica, por su parte, es el estudio de los signos y códigos en los sistemas vivientes. Una noción fundamental en este ámbito es que «la existencia del código genético implica que toda célula es un sistema semiótico». ¹⁵ Barbieri afirma que la biología moderna no ha aceptado esta premisa fundamental de la biosemiótica debido a que en la primera hay tres conceptos básicos que son incompatibles con ella. El primero es la descripción de la célula como una especie de ordenador. En esta metáfora se considera que los genes (la información biológica) constituyen el software, y las proteínas el hardware. Los ordenadores tienen códigos, pero no son sistemas semióticos, ya que dichos códigos provienen del exterior del sistema a través de un codificador externo, mientras que, como acabamos de ver, un sistema semiótico «incluye» al propio codificador. El concepto de la célula como ordenador también implica la idea de que el código genético surgió de un codificador ajeno al sistema: la selección natural. Según esta descripción, los seres vivos no son sistemas semióticos y el «código genético» no es más que una metáfora.

El segundo concepto en cuestión entre la biología moderna y la biosemiótica es el fisicalismo, esto es, la noción de que todo es reducible a cantidades físicas. Los biólogos exigen que todas las cosas (ADN, moléculas, células, organismos...) obedezcan a leyes que determinan su comportamiento. Un código semiótico se caracteriza por tener el vacilante aspecto propio de las reglas, y no por regirse por leyes físicas deterministas que vinculen inexorablemente los símbolos con su significado. Por último, la tercera fuente

de discordia es la convicción (o la falta de ella) de que toda innovación biológica es resultado de la selección natural.

Barbieri sostiene que los biólogos están pasando por alto algo muy básico al establecer esos presupuestos fundamentales: ignoran el origen de la vida. La evolución por selección natural requiere la copia de registros genéticos y la construcción de proteínas, pero esos procesos en sí tuvieron que originarse de alguna manera. Barbieri señala que los genes y las proteínas de los sistemas vivientes son distintos de todas las demás moléculas, principalmente porque se producen de forma completamente diferente.

La estructura de las moléculas del mundo inorgánico, el mundo de los objetos como los ordenadores y las rocas, viene determinada por los enlaces que se forman de manera espontánea entre sus átomos. Dichos enlaces vienen determinados a su vez por factores internos, las características químicas y físicas inherentes a los átomos. Felizmente deterministas.

No ocurre así en los sistemas vivientes. Los genes son complejas secuencias de nucleótidos, mientras que las proteínas son complejas secuencias de aminoácidos. Esas secuencias no se juntan espontáneamente en una célula. No hay un amor a primera vista que las lleve a unirse atraídas por una química irresistible. Lejos de ello, se agrupan de manera improvisada gracias a la acción de una clase específica de moléculas, de todo un sistema integrado por ácido ribonucleico (ARN) y «casamenteras» proteínicas que las ayudan a hacerlo. Barbieri señala que esto tiene una gran importancia por sus implicaciones en relación con el origen de la vida.

Las primitivas moléculas «enlazadoras» —las precursoras del sistema del ARN, que une los nucleótidos— surgieron mucho antes de que lo hicieran las primeras células. Lo mismo ocurrió con ciertas moléculas enlazadoras que desarrollaron la capacidad de unir nucleótidos siguiendo una plantilla: las «copiadoras». Esas enlazadoras y copiadoras empezaron a existir en función de recursos moleculares aleatorios. Y fue la existencia de moléculas copiadoras la que desencadenó el proceso de la evolución. La selección

natural cinceló la existencia de los seres vivos, pero las moléculas necesarias para la evolución —las enlazadoras y copiadoras— existían antes que la vida misma.

Barbieri afirma que «la selección natural es el resultado a largo plazo de la copia molecular, y sería el único mecanismo evolutivo si la copia fuera el único mecanismo básico de la vida». ¹⁶ Pero no lo es. Aunque los genes pueden ser su propia plantilla y copiarse a sí mismos, las proteínas no: estas no pueden fabricarse copiando otras proteínas. El problema es que solo las moléculas que se pueden copiar pueden heredarse, de modo que la información acerca de cómo fabricar las proteínas tiene que provenir de los genes. Barbieri señala que la característica sobresaliente de los primeros fabricantes de proteínas «fue la capacidad de asegurar una correspondencia específica entre genes y proteínas, porque sin ella no habría especificidad biológica, y sin especificidad no habría herencia ni reproducción. La vida tal como la conocemos simplemente no existiría sin una correspondencia específica entre genes y proteínas». ¹⁷ La correspondencia específica a la que se refiere es un código. Y ese código tenía que estar allí primero, antes de que se pusiera en marcha la selección natural.

Esto es lo interesante para nuestro propósito: si esa correspondencia no fuera un código, sino que viniera determinada por la estereoquímica —que es lo que se suponía inicialmente—, sería automática y, por ende, determinista. Pero no es ese el mecanismo, lo cual constituyó una auténtica sorpresa para los biólogos. El puente entre los genes y las secuencias de aminoácidos que codifican y que componen las proteínas lo proporcionan las moléculas del ARN de transferencia. Dichas moléculas tienen dos sitios de reconocimiento distintos: uno es para un codón (un grupo de tres nucleótidos), y el otro para un aminoácido, uniéndose así a ambos elementos. También esta podría ser la configuración de una correspondencia automática entre un codón y un aminoácido concreto si un sitio de reconocimiento determinara físicamente qué se une al otro; pero no es así. Los dos sitios son independientes uno de

otro y además están físicamente separados. Señala Barbieri: «Simplemente no existe un vínculo necesario entre los codones y los aminoácidos, y la existencia de una determinada correspondencia entre ellos solo puede ser el resultado de reglas convencionales. Solo un verdadero código, en suma, podía garantizar la especificidad biológica, y eso significa que de ninguna manera puede descartarse el código genético como una mera metáfora lingüística». Lo que nos lleva a la siguiente conclusión: «La célula es un verdadero sistema semiótico porque contiene todas las características esenciales de tales sistemas, es decir, signos, significados y código, todo ello producido por el mismo codificador». ¹⁸

Las evidencias que respaldan este concepto de sistema biosemiótico y contradicen los conceptos fundamentales de la biología moderna son extremadamente recientes. Hace poco los científicos han descubierto que los cefalópodos (la familia que incluye al pulpo) pueden recodificar su ARN. Las moléculas de ARN tienen el privilegio de establecer códigos con ADN (en la parte del ARN que reconoce la secuencia del codón de tres nucleótidos del ADN) y también con proteínas (en la parte del ARN —recordemos que distinta de la anterior— que reconoce el aminoácido). Recodificar el ARN implica que pueden construirse nuevas proteínas mientras que la secuencia de símbolos del ADN permanece igual. El resultado colectivo es la destrucción de la correspondencia unívoca entre genes y proteínas. La recodificación permite que un solo gen de pulpo produzca muchos tipos diferentes de proteínas a partir de la misma secuencia de ADN. ¹⁹ Esto tiene una gran importancia, ya que constituye una evidencia en contra de los tres conceptos de la biología que rechazan la noción de sistemas semióticos en los organismos vivos. El sistema puede modificar su código; posee un codificador interno capaz de producir innovaciones biológicas (nuevas proteínas), pero no a través de la selección natural. Esto ilustra la arbitrariedad de la conexión de un símbolo con su significado en un sistema viviente.

Si en los sistemas vivientes los símbolos son arbitrarios y el ARN es el

codificador, ¿qué pinta aquí el ADN? ¿Por qué este último ha tenido el monopolio del simbolismo molecular durante los últimos cientos de millones de años? En su manifestación física, y a diferencia del ARN, el ADN es extremadamente estable a nivel estructural, lo que le ha ayudado a ser siempre la estructura simbólica preferida a lo largo de toda la evolución. Sin embargo, aunque actualmente el ADN de nuestras células y de las células de otros organismos vivientes es muy estable, en el origen de la vida su estructura no tenía esa cualidad. La mezcla y la reorganización aleatoria de moléculas, a través del proceso irreversible y probabilístico de la selección natural, generó moléculas similares a las bases de nucleótidos. Mediante posteriores procesos de reorganización, los componentes y secuencias de ADN de mayor éxito sobrevivieron y se replicaron.

Pero ¿qué entendemos por «éxito» cuando hablamos del ADN? El ADN está compuesto por cuatro nucleótidos distintos. Los genes son secuencias de combinaciones concretas de nucleótidos que actúan como la descripción simbólica —la receta— para fabricar proteínas. ¿Qué implica el éxito de una secuencia de ADN? ¿Hablamos de la capacidad de mantenerse físicamente estable durante toda la vida de un organismo? ¿O nos referimos más bien a su fiabilidad a la hora de codificar la información necesaria para la replicación fructífera del organismo? Pues ambas cosas. El ADN —o, mejor dicho, la estructura de la memoria hereditaria cuando constriñe la construcción del ADN— obedece a las leyes de Newton y se mantiene termodinámicamente estable en el entorno acuoso de la célula debido a las propiedades de sus bases de nucleótidos. Sin embargo, en su modo informativo (subjetivo) el ADN sigue reglas, no las leyes de la física. Las secuencias de bases que sobreviven han sido seleccionadas por la evolución según una regla: elegir la información más fiable y útil para la supervivencia y la reproducción del organismo. Los nucleótidos que componen el ADN y transportan información en forma simbólica son el resultado de una selección, y, aunque son arbitrarios, se han conservado en una forma estable durante todo el proceso de

la evolución por haber hecho un buen trabajo y por seguir haciéndolo, a diferencia de los profesores universitarios que aflojan la marcha tras conseguir el puesto que anhelaban.

Durante la replicación, esos nucleótidos se leen y traducen en secuencias lineales de aminoácidos (que forman enzimas y proteínas) mediante un proceso regido por reglas. Ese conjunto de reglas es lo que se conoce como «código genético». El ADN contiene la secuencia, pero las que ejecutan el código son moléculas de ARN. Ciertas secuencias de ADN, llamadas codones —que, como ya hemos visto, están formadas por tres nucleótidos—, simbolizan ciertas secuencias de aminoácidos. No hay ambigüedad, pero tampoco hay solo un codón por cada aminoácido. Por ejemplo, se requieren seis codones distintos para simbolizar la arginina, pero solo uno para simbolizar el triptófano. Sin embargo, los componentes de la secuencia de ADN (el símbolo) no se asemejan a los componentes de la cadena de aminoácidos (su significado), del mismo modo que las palabras que simbolizan los componentes de una receta no se parecen en nada a los componentes propiamente dichos.

Cuando una secuencia de ADN se ha traducido en una cadena de aminoácidos, la actividad instructora del ADN llega a su fin (aunque solo temporalmente). No ocurre así, en cambio, con las restricciones que esa cadena de símbolos ha impuesto a la estructura material de dichos aminoácidos. Una vez construida la cadena (recuérdese que los aminoácidos no se unen entre sí de manera espontánea), esta se pliega sobre sí misma, formando enlaces químicos débiles entre las moléculas que actúan casi como débiles imanes. Qué enlaces se forman y cómo se pliega la cadena depende de la posición que ocupe cada aminoácido, la cual viene dictada por la descripción simbólica. Esta es la parte difícil. Una vez situados esos aminoácidos, los enlaces que se forman vienen determinados por leyes físicas. Ciertos aminoácidos huyen del agua, mientras que a otros les gusta; ciertos aminoácidos se adhieren unos a otros, a veces de forma bastante intensa. La

interacción de la cadena de aminoácidos con su entorno pliega la cadena en una estructura tridimensional, una proteína.²⁰ Este plegado es lo que transforma la secuencia de aminoácidos, que es lineal, unidimensional y regida por reglas, en una estructura de control funcional dinámica, tridimensional y regida por leyes (una proteína).

Obviamente, las proteínas obedecen a las leyes causales de la física y la química, pero es la información simbólica arbitraria de las secuencias de ADN la que determina tanto la composición material como la función bioquímica de las proteínas. Resulta bastante impresionante. El ADN constituye el ejemplo primigenio de información simbólica (las secuencias de nucleótidos) que controla una función material (la acción de las enzimas) ligada a un código regido por reglas, tal como Von Neumann había predicho que debía ocurrir en el caso de los autómatas dotados de la capacidad de autorreproducirse y de evolucionar. Pero aguarde un momento: ¿qué fabricó la proteína? El ADN tenía la información decodificada necesaria para fabricarla, pero ¿qué fue lo que desencadenó el proceso? La respuesta es: otra proteína. Hacía falta la intervención de una enzima (una proteína) para desgarrar las hebras del ADN y poner en marcha todo el proceso de replicación. De hecho, fue una enzima recién creada la que separó dichas hebras. Es el viejo problema del huevo y la gallina: sin enzimas catalizadoras que separen las hebras del ADN, este no pasa de ser un mensaje inerte que no puede replicarse, transcribirse ni traducirse; pero sin ADN no habría enzimas catalizadoras. Es la complementariedad de Bohr: dos partes complementarias, dos modos de descripción, que conforman un solo sistema.

Von Neumann, en la exposición de su experimento mental sobre la autorreplicación, escribió que había evitado la pregunta «más intrigante, emocionante e importante de por qué las moléculas o agregados que realmente se dan en la naturaleza [...] son lo que son, por qué son esencialmente moléculas muy grandes en algunos casos, pero grandes agregados en otros». ²¹ Pattee sugería que es el propio tamaño de las moléculas el que une

los mundos cuántico y clásico: «Las enzimas son lo bastante pequeñas para aprovechar la coherencia cuántica a fin de obtener el enorme poder catalizador del que depende la vida, pero lo bastante grandes para lograr una alta especificidad y arbitrariedad en la generación de productos eficazmente decoherentes capaces de *funcionar* como estructuras clásicas». ²² La decoherencia cuántica significa básicamente que las partículas subatómicas se sincronizan para «cooperar» con el fin de generar productos decoherentes, que son partículas que no tienen propiedades cuánticas. Pattee señala que en la actualidad hay investigaciones que respaldan su hipótesis de que las enzimas requieren efectos cuánticos, ²³ pero, a la vez, en un mundo estrictamente cuántico la vida sería imposible. ²⁴ Ambas son necesarias: una capa cuántica y una capa física clásica.

La serpiente que se muerde la cola: cierre semiótico

Von Neumann dejó claro que su autómeta necesitaría replicarse. Para poder autorreplicarse hay que especificar los límites del yo. Para hacer un «yo» se necesitan piezas que implementen su descripción, traducción y construcción. Y para que este yo haga otro yo es necesario describir, traducir y construir piezas que describan, traduzcan y construyan. Este bucle autorreferencial no es un mero quebradero de cabeza: se trata de lo que se conoce como «cierre lógico», que en este caso define un «yo».

Pattee denomina «cierre semiótico» al conjunto de condiciones físicas necesarias para que se dé esta interdependencia excepcional de símbolo-materia-función, y subraya que, para ejecutar físicamente este cierre, las instrucciones simbólicas han de tener una estructura material. No puede haber fantasmas en el sistema, y la estructura física debe constreñir todos los procesos dinámicos de construcción legítimos según las leyes de Newton. El cierre del bucle semiótico, la unión física de las moléculas, es lo que define los límites del «yo», el sujeto, en proceso de «autorreplicación». No se

incorpora ninguna estructura aleatoria que ande por ahí suelta; se han establecido los límites. Eso no implica que la célula tenga alguna forma de autopercepción; pero tampoco puede haber autopercepción sin un yo. Los primeros pasos deben dirigirse, pues, hacia la delimitación del yo. Los siguientes destinos, la autopercepción, el autocontrol, la experiencia de sí, la conciencia de sí y la capacidad de ensimismamiento, están todos ellos mucho más adelante en el camino.

El cierre semiótico debe estar presente en todas las células que se autorreplican. Es cierto que «el yo» se ha ido haciendo más elaborado a través de los procesos evolutivos, pero hasta una célula sigue el consejo de Harry el Sucio y «conoce sus limitaciones». Cualesquiera procesos físicos complejos que cierren el bucle símbolo-materia constituyen el puente que salva el *Schnitt* del físico, la brecha explicativa, el abismo entre sujeto y objeto; forman el protocolo entre la capa cuántica y la newtoniana. Los procesos que cierran el bucle símbolo-materia unen los dos modos de descripción, salvando la brecha que surgió en el origen de la vida, lo cual implica que, de la misma forma, la brecha entre la experiencia consciente subjetiva y los estímulos neuronales objetivos de nuestro cerebro físico, entre estos dos modos de descripción, puede salvarse mediante un conjunto similar de procesos, e incluso es posible que dichos procesos se estén produciendo en el interior de las células.

La rendición y la tregua

En los primeros días de la física cuántica, Niels Bohr presentó el principio de complementariedad como una bandera blanca, un intento de explicar la doble naturaleza de la luz (su dualidad onda/partícula). El principio de complementariedad considera que tanto las leyes causales objetivas como las reglas de medición subjetivas son fundamentales para la explicación de los fenómenos. Bohr subrayó que, si bien se necesitaban dos modos de descripción, ese hecho no se correspondía con una dualidad real del sistema en observación. El sistema en sí estaba unificado. Era las dos cosas a la vez; las dos caras de una misma moneda.

Eso es lo que hace que nos resulte tan difícil de entender, si es que llegamos a hacerlo. De hecho, Richard Feynman decía: «Creo que puedo afirmar con seguridad que nadie entiende la mecánica cuántica». Bohr hacía una analogía con la distinción entre sujeto y objeto, que amplió a la mente y la materia en su conferencia de 1927 en Como: «Espero [...] que la idea de complementariedad resulte adecuada para describir la situación, que guarda una profunda analogía con la dificultad general en la formación de las ideas humanas inherente a la distinción entre sujeto y objeto». ²⁵

Pattee, en cambio, se muestra más audaz. Él ve algo más que una mera analogía: ve la complementariedad como una necesidad epistemológica que surgió en el origen de la vida y se extiende a todos los niveles evolutivos. Su esencia no es meramente el reconocimiento de la división sujeto/objeto, sino «la articulación aparentemente paradójica de los dos modos de conocimiento». ²⁶ Esa paradoja ha sumido a los filósofos y científicos en una disputa dualista durante más de dos mil años; y, si siguen por el camino actual, seguirán peleándose durante otros dos mil. Los dos modos de investigación, los dos fenómenos que estos han revelado, no se describen mediante el mismo conjunto de leyes físicas. Pattee no puede por menos que dejar escapar una sonrisa burlona ante el hecho de que el modo objetivo ha llevado a los «reduccionistas a afirmar que la vida no es más que física ordinaria, que de hecho lo es mientras uno no esté dispuesto a considerar los problemas subjetivos de las mediciones y las descripciones [...] Lo que dice el principio de complementariedad es que usando solo este modo de descripción objetivo ni siquiera la física es reducible a él». ²⁷

Al igual que los científicos y los filósofos tuvieron que aceptar el hecho de que el mundo no era plano, vamos a tener que lidiar con el principio de complementariedad aplicado a la mente y el cerebro. Dicho principio todavía resulta controvertido porque choca frontalmente con la creencia de que la mejor explicación de algo es una explicación única. Pero en física la falacia de la explicación única se extinguió hace cien años con el descubrimiento del

mundo cuántico. El mundo microscópico sigue leyes distintas de las del mundo macroscópico; uno y otro habitan en diferentes capas de descripción, y no son mutuamente reducibles.

Quienes se adhieren al «patrón oro» de la explicación única simplemente ignoran las realidades de la física. Pattee lamenta que «la aceptación [de la complementariedad] en la mecánica cuántica solo se produjera tras el fracaso de cualquier otra interpretación». ²⁸ Esto recuerda a la célebre frase de Sherlock Holmes: «Cuando descartas lo imposible, lo que queda, por improbable que parezca, debe de ser la verdad». Pattee se pregunta si a la aceptación de la complementariedad en las teorías biológicas y sociales le aguarda el mismo angustioso destino. Como dijo bromeando Richard Feynman en cierta ocasión: «Si no te gusta vete a otro sitio. [...] Vete a otro universo donde las reglas sean más simples, y la filosofía más agradable, más psicológicamente fácil». ²⁹ Que no te guste la idea no significa que las cosas no sean así.

Resumiendo

La materia viva es distinta de la inanimada porque ha tomado un rumbo completamente distinto. La materia inanimada obedece a leyes físicas. La vida, desde el principio, ha apostado por las reglas, los códigos y la arbitrariedad de la información simbólica. La distinción entre la información simbólica y la materia, así como la interdependencia entre ambas, han posibilitado una evolución abierta que ha dado como resultado la vida tal como la conocemos. La información relativa a las situaciones de éxito del pasado se almacenó en registros integrados por símbolos. Dichos registros son en sí mismos mediciones, cuya naturaleza es intrínsecamente probabilística. Sin embargo, la vida depende de esos símbolos arbitrarios y probabilísticos para su propia construcción material en el mundo físico. La arbitrariedad inherente a los símbolos y las mediciones proporciona cierta «salsa», es decir,

algún elemento de imprevisibilidad, que se combina con el predecible «menú» de las leyes físicas, lo que hace que la vida se vuelva cada vez más ordenada y cada vez más compleja a lo largo del tiempo.

Esa distinción entre sujeto y objeto no es solo una interesante rareza. Se inicia en el nivel de la física en la distinción entre la probabilidad inherente a las mediciones simbólicas y la certeza de las leyes materiales. Más tarde la distinción se ejemplifica en la diferencia entre el genotipo, la secuencia de símbolos-nucleótidos que componen el ADN de un organismo, y el fenotipo, su estructura física real tal como la prescriben dichos símbolos. Y viaja con nosotros a través de las diversas capas evolutivas hasta la distinción entre la mente y el cerebro.

Durante los últimos veinticinco años, los debates sobre el pensamiento y la conciencia se han centrado en el ser humano y, más recientemente, en el cerebro humano plenamente evolucionado, pero eso no nos ha ayudado a avanzar para salvar la brecha explicativa. Es hora de que empecemos a explorar la brecha planteada por Howard Pattee entre la materia viva y la inanimada. Si logramos determinar cómo se salvó, cómo la vida logró efectuar el cierre semiótico, quizá podamos entender cómo salvar también la brecha explicativa entre mente y cerebro. En esto contamos incluso con el apoyo de William James, que llegó a considerar la posibilidad de lo que él denominaba la teoría del «polizoísmo»: «Cada célula cerebral tiene su propia conciencia individual, de la que ninguna otra célula sabe nada, dado que todas las conciencias individuales son mutuamente “ejectivas”». ³⁰ Para él, la célula individual alberga un proceso muy rudimentario que conecta un «yo» subjetivo con la mecánica objetiva. El cierre semiótico, el vínculo que salva la brecha entre la materia viviente y la no viviente, está presente en todas las células. Asimilando este hecho y tratando de entender los procesos en él involucrados, podemos empezar a buscar una interpretación de la conciencia que parta desde una perspectiva distinta, y a buscarla en diferentes lugares. No estoy sugiriendo con ello que las células individuales sean conscientes, sino que

pueden tener algún tipo de procesamiento que sea necesario o similar al procesamiento que se traduce en la experiencia consciente.

La brecha explicativa nos ha dejado perplejos porque las experiencias subjetivas de la mente se han resistido a dejarse reducir a los estímulos neuronales de la materia cerebral. Unas y otros parecen ser dos propiedades complementarias irreductibles de un sistema único. Sabemos que, por mucho que aprendan los observadores externos objetivos sobre la estructura, la función, las actividades y los estímulos neuronales del cerebro, la experiencia que tiene el propio sujeto de dichos estímulos es completamente distinta de cualquier observación de ellos. Los detalles de la estimulación neuronal, o incluso el mero hecho de que haya neuronas estimuladas, no son parte de la experiencia ni de las intuiciones del sujeto. El funcionamiento objetivo de la percepción, el pensamiento, etc., no resulta visible para la persona que percibe o que piensa. Como veíamos en el capítulo sobre las capas, esos detalles no son necesarios para la persona y, por lo tanto, están ocultos, abstraídos de la vista. Asimismo, la función de las neuronas no puede derivarse de su estructura sin ningún conocimiento previo de dicha función, ni su estructura puede derivarse de esta última. Saberlo todo sobre una de ellas no nos dice nada de la otra. Son dos capas separadas e irreductibles, con protocolos distintos. Pattee cree que esto es parte integrante del principio de complementariedad, y que un modelo único no puede explicar tanto la estructura objetiva como la función subjetiva. El corte epistémico, el corte sujeto/objeto, está vivo y coleando en el nivel del cerebro humano. Pattee sostiene que «nuestros modelos de organismos vivientes nunca eliminarán la distinción entre el yo y el universo, porque la vida empezó con esa diferenciación, y la evolución la requiere». ³¹

Así pues, no debería sorprendernos que sigan apareciendo en nuestro pensamiento dos modos de comportamiento complementarios, dos niveles de descripción. El corte sujeto/objeto está presente en todos los grandes debates filosóficos: aleatorio/predecible, experiencia/observación, individuo/grupo,

cultura/natura y mente/cerebro. Pattee considera que estos dos modos complementarios son inevitables y necesarios para cualquier explicación que vincule los modelos de experiencia subjetivo y objetivo. Ambos modelos son inherentes a la propia vida, han estado presentes desde el principio y han sido preservados por la evolución. Escribe Pattee: «Se trata de una complementariedad universal e irreductible. Ninguno de los modelos puede derivarse del otro ni reducirse al otro. Por la misma lógica por la que un modelo objetivo detallado de un dispositivo de medición no puede producir la medición de un sujeto, tampoco un modelo objetivo detallado de un cerebro material puede producir el pensamiento de un sujeto». ³²

Ignorar uno de los dos lados de la brecha hará que se pierda el vínculo entre ellos. Y vincular a ambos requiere reconocer la naturaleza dual y complementaria de los símbolos. El vínculo estará integrado por mecanismos que la física puede describir, pero es posible que la explicación no resulte cálida y reconfortante, y que no satisfaga psicológicamente a nadie, ni a los deterministas ni a quienes creen en los espíritus. Puede ser, como ocurre con la mecánica cuántica, algo que nadie acabe de entender del todo, que esté mucho más allá del alcance de nuestras intuiciones e imaginaciones. Afirmaba Feynman con tono reprensor: «Nosotros no vamos a decirle a la naturaleza lo que tiene que ser. Eso es lo que hemos descubierto. Cada vez que tratamos de adivinar cómo tiene que ser y nos ponemos a medirlo, ella revela su inteligencia. Siempre es más imaginativa que nosotros y encuentra una forma más inteligente de hacerlo de la que nosotros habíamos pensado». ³³

Torrentes de burbujas y conciencia personal

Sería estupendo que para variar algo tuviera sentido.

ALICIA, EN *Alicia en el País de las Maravillas*,
DE LEWIS CARROLL

Todos compartimos ese estado que llamamos conciencia, esa percepción del flujo de nuestros propios pensamientos, anhelos, emociones y sentimientos sobre el mundo, sobre el prójimo y sobre nosotros mismos. No solo es omnipresente, sino que además tiene un carácter personal, definitorio y delimitador. Define la experiencia de vivir. El yo consciente parece estar por encima del cerebro físico y de todas sus capas y módulos. Parece que sin él seríamos como los autómatas que Descartes observara en los jardines parisinos: máquinas. Siendo así, ¿qué características podría tener una posible explicación de la conciencia?

Como podrá usted imaginar, las piezas del rompecabezas que en mi opinión nos pueden conducir a una nueva concepción sobre la naturaleza de la experiencia consciente son las que hemos abordado en los capítulos anteriores: los módulos, las capas, el principio de complementariedad y el cierre semiótico de Howard Pattee. Estos conceptos nos ayudarán a entender que los circuitos neuronales son estructuras con una doble vida: transportan información simbólica, que está sujeta a reglas arbitrarias, pero a la vez poseen una estructura material sometida a las leyes de la física. Estas dos perspectivas combinadas nos cuentan la historia del cerebro, un órgano

ingeniosamente diseñado por la selección natural y organizado en módulos locales cuyo funcionamiento tiene lugar en una arquitectura estratificada, de modo que, en general, un conjunto de módulos no sabe lo que hacen los demás. Es la historia de un puñado de pequeños y esforzados circuitos enmarcados en una organización coherente para producir una función de mayor envergadura; del mismo modo que los ciudadanos, dada su inteligencia, trabajan en relativo aislamiento para producir uno u otro tipo de sociedad. El secreto para comprender su funcionamiento debe incluir el modo como esas esforzadas piezas se expresan en cada momento.

Si las brechas, módulos y capas han de ayudarnos a entender cómo los cerebros crean mentes, también tienen que dar cuenta de algunos hechos persistentes relativos al cerebro. Dedicemos un momento a considerar detalladamente esta cuestión y sus implicaciones para nuestra percepción del yo: un neurocirujano puede desconectar los dos hemisferios del cerebro y producir dos mentes en una misma cabeza, dos mentes con diferentes contenidos a la vez, pero con los mismos impulsos y sentimientos emocionales. Recordemos también esto: mientras que una lesión cerebral puede causar déficits específicos, resulta casi imposible eliminar la conciencia por completo. Y por último: si bien la experiencia consciente parece algo unificado y completo, en realidad tiene lugar de una forma concertada, con múltiples sistemas que discurren paralelos entre sí, cada uno de los cuales arroja los resultados de su procesamiento por separado. Así pues, aunque la conciencia nos parezca una película coherente y perfectamente editada, en realidad es más bien un flujo de fotogramas individuales que emergen como agitadas burbujas en una olla de agua hirviendo, vinculados solo por su sucesiva aparición en el tiempo. La conciencia se halla en constante cambio, es como un torrente; y como dijo en cierta ocasión William James: «Ningún estado, una vez desaparecido, puede reaparecer y ser idéntico al que había sido antes».¹ Permítame preparar el terreno para esta idea.

Dos hemisferios conscientes, conscientemente distintos

Volvamos a la que sería la primera de mis observaciones científicas. Fue la del «caso W. J.», un hombre que sufría ataques epilépticos tan fuertes que solo podía llevar una vida normal dos días a la semana. Un joven neurocirujano llamado Joseph Bogen, que había realizado una extensa investigación, sugirió que W. J. podría beneficiarse de una operación quirúrgica raras veces realizada en la que se cortaba el importante tramo de nervios que conecta los dos hemisferios cerebrales. La operación se había hecho ya veinte años antes con varios pacientes en Rochester, Nueva York, con el fin de controlar la epilepsia, y se había logrado reducir los ataques de forma drástica. Curiosamente, después de ver reducido su cerebro a la mitad, todos los pacientes dijeron que se sentían bien, y la única diferencia que notaron fue la desaparición de los ataques.

W. J. era un veterano de la Segunda Guerra Mundial que había combatido en muchas batallas. Tras sopesar las probabilidades de librar esta, aceptó la cirugía. Yo era el joven estudiante de posgrado encargado de diseñar las pruebas que había que hacerle tras someterle a la operación del cerebro escindido para ver qué efectos —en el caso de haberlos— había provocado la cirugía en su función cerebral. Se esperaba que no hubiera ninguno, dado que tampoco se había detectado efecto alguno en los pacientes de Rochester. W. J. era un hombre afable y cordial, y los dos hemisferios de su cerebro parecían funcionar muy bien juntos a pesar de no estar ya en comunicación directa. Uno de ellos hablaba; el otro no. Dada la configuración neuronal del cerebro —su cableado—, eso significaba que el hemisferio izquierdo, el «parlante», veía el mundo situado visualmente a la derecha de un punto fijo, mientras que el derecho, el «no parlante», veía toda la información visual situada a la izquierda del mismo punto. Dado ese estado quirúrgico, me pregunté: si proyectara una luz hacia la derecha, ¿W. J. diría que la había visto? La luz debería ir al hemisferio izquierdo, y el hemisferio izquierdo estaba dotado del

habla; debería ser fácil. En efecto lo fue, y W. J. declaró sin problemas haberla visto.

Poco después proyecté la misma luz hacia el lado izquierdo del espacio y esperé a ver si me decía algo. No lo hizo. Entonces le pregunté directamente si había visto alguna cosa, y la respuesta fue un contundente «No». ¿Acaso estaba ciego de ese lado? ¿O era simplemente que ese punto de luz ya no se comunicaba con el hemisferio que hablaba? ¿Sabía el hemisferio derecho, aparentemente mudo, que había visto una luz? ¿Era consciente de ella? ¿Qué estaba ocurriendo?

Fue más avanzada la sesión de pruebas cuando se hizo evidente que el hemisferio derecho de W. J había visto la luz, puesto que fue capaz de apuntar de manera fácil y precisa hacia ella con la mano izquierda. Aquella fue la primera de una serie de observaciones que revelarían que no solo se había dividido el cerebro, sino también la mente; la semilla que conduciría a sesenta años de investigaciones sobre la naturaleza de la mente y sus fundamentos físicos. Y también fue la prueba que produjo la observación más asombrosa de todas: el hemisferio izquierdo no parecía echar de menos al derecho y viceversa. No era solo que no lo echara de menos, es que ni siquiera lo recordaba; como tampoco recordaba las funciones que había realizado: era como si el otro hemisferio no hubiera existido nunca. Para mí, este fenómeno constituye el dato más importante que debe tener en cuenta cualquier estudioso que investigue la relación mente/cerebro.

¿Cómo es que el hemisferio izquierdo no se queja por el hecho de no ser ya consciente de las cosas situadas en el lado izquierdo del espacio? Imagine que le desconectan los dos hemisferios del cerebro. Imagine que se despierta en una habitación de hospital a la mañana siguiente y, cuando entra su cirujano para ver qué tal está, usted solo ve la mitad izquierda de su rostro. ¿No cree que notaría la ausencia de la mitad derecha? Pues resulta que no. De hecho, su hemisferio izquierdo ni siquiera sería consciente de que hay una mitad izquierda del espacio. Pero aquí viene la parte más extraña. Yo he hablado

como si la nueva versión dividida de usted fuera solo su hemisferio izquierdo, y eso no es así: usted también es su hemisferio derecho. Los nuevos «ustedes» tienen dos mentes, con dos acervos completamente independientes de información perceptual y cognitiva; con la única diferencia de que solo una de las dos mentes puede hablar con facilidad. La otra en principio no puede; quizá, después de muchos años, llegue a generar algunas palabras.

Un hecho aparentemente aún más descabellado es que en los primeros meses después de la cirugía, antes de que los dos hemisferios se acostumbren a compartir un mismo cuerpo, se los puede ver enzarzados en una especie de tira y afloja. Por ejemplo, hay una sencilla prueba en la que uno debe organizar un pequeño conjunto de bloques de colores para que coincidan con un patrón que se muestra en una tarjeta. El hemisferio derecho alberga las especializaciones visomotoras que hacen que esta prueba sea un juego de niños para la mano izquierda, mientras que el izquierdo, por su parte, es del todo incompetente para este tipo de tareas. Cuando un paciente cuyo cerebro ha sido escindido recientemente intenta realizar la tarea, la mano izquierda resuelve el rompecabezas de inmediato; pero cuando lo intenta con la derecha, la mano izquierda empieza a entorpecer el trabajo de esta, tratando de entrometerse y de completar ella la tarea. En una de esas pruebas tuvimos que hacer que el paciente se sentara sobre su autoritaria mano izquierda para permitir que la derecha intentara realizar la tarea, que no logró completar: estaba más allá de las capacidades del hemisferio izquierdo.

Cuando se pierde la comunicación entre los dos hemisferios, cada uno de ellos deja de ser consciente de los conocimientos del otro y actúa de manera independiente en función de la información que recibe. Ambos lados del cerebro intentan realizar la tarea por separado, lo que se traduce en un tira y afloja entre ellos. La sencilla tarea mencionada revela el espejismo de una conciencia unificada: es obvio que, si la conciencia surgiera de un único lugar, el paciente de cerebro escindido no podría tener dos experiencias simultáneas.

Pero las cosas todavía se ponen más interesantes. Todos hemos visto la

película en la que se muestra la sencilla ilusión óptica que se produce cuando dos bolas parecen chocar entre sí, y, tras la falsa colisión, la pelota que supuestamente recibe el impacto sale disparada. En la jerga psicológica, esto se conoce como «efecto de lanzamiento de Michotte», en honor al psicólogo belga Albert Michotte, que fue quien diseñó la ilusión para investigar cómo percibimos e inferimos la causalidad. Dado que la primera bola se detiene junto a la segunda sin llegar a tocarla, en realidad no se produce ningún acontecimiento físico que haga que esta última salga disparada. Pero no es así como todos percibimos la experiencia: la bola A golpea la bola B y esta sale disparada; punto. Es decir, que ha habido causalidad.

Pero ¿cómo ven esta ilusión óptica los pacientes con cerebro escindido? ¿La ve el hemisferio izquierdo, a través del prisma de la conciencia que posee, tal como podría verla el derecho? Uno de los primeros experimentos diseñados para comprobarlo fue ideado por Matthew Roser, un estudiante de Nueva Zelanda que por entonces trabajaba en mi laboratorio de Dartmouth y actualmente reside en la Universidad de Plymouth, en el Reino Unido.² Matt, un científico excepcionalmente dotado, examinó junto con otros colegas cómo cada hemisferio, trabajando solo y desconectado del otro, veía las bolas que parecían colisionar. Los resultados fueron sorprendentes: el hemisferio derecho captaba instantáneamente la ilusión, mientras que el izquierdo no. Este hecho quedó establecido mediante un segundo experimento en el que se incrementaba un poco la distancia entre las bolas en el punto de la falsa colisión, o bien se aumentaba ligeramente el tiempo que tardaba la segunda en moverse. En estas circunstancias, para el hemisferio derecho la ilusión desaparece, es decir, que no llega a producirse. En cambio el hemisferio izquierdo, que es el que realiza el trabajo pesado en todos los asuntos cognitivos, se mostraba incapaz de detectarla en ninguna circunstancia. Curiosamente, el hemisferio izquierdo detecta asociaciones que el derecho parece incapaz de captar. En el contexto de estas pruebas, fue el cerebro izquierdo el que resolvió otro problema, que requería de inferencia lógica, y

que el derecho no supo solventar. En suma: la «percepción» directa de la causalidad estaba al alcance del hemisferio derecho, pero la capacidad para «inferir» dicha causalidad solo formaba parte de la caja de herramientas del izquierdo.

Cuando examinamos cómo un cerebro normalmente conectado percibe estos dos tipos de tareas, de inmediato se hace patente que ante una ilusión óptica es el hemisferio derecho el que dispone del equipamiento neuronal necesario para detectarla, mientras que a la hora de abordar una tarea de inferencia lógica es el izquierdo el que procesa la información. Así pues, en un cerebro conectado, en el momento A, cuando el hemisferio derecho ve el falso lanzamiento de la bola, dice: «¡Eh!, la bola A acaba de golpear a la bola B»; pero en el momento B, cuando el cerebro se enfrenta a una prueba de inferencia lógica, es el hemisferio izquierdo, no el derecho, el que se revela capaz de detectarla. Es una situación similar a algunos de los juegos que pueden verse en los salones recreativos. Percibimos la información procesada —es decir, somos conscientes de ella— a medida que esta emerge después de procesarse en un hemisferio concreto. Pero ¿eso ocurre porque todo proceso neuronal activa una red de «hazlo consciente» (que debería estar presente en ambos hemisferios) o es que cada proceso posee en sí mismo y por sí solo la capacidad neuronal de hacerse consciente?

Diminutas burbujas

Yo me inclino por lo segundo. Pensar en esta cuestión me llevó a emplear la metáfora del agua burbujeante como una forma de conceptualizar cómo se despliega nuestra conciencia. No es que esta sea el producto de una red especial que permite que todos nuestros acontecimientos mentales sean conscientes; lejos de ello, cada acontecimiento mental es gestionado por módulos cerebrales que poseen la capacidad de hacernos conscientes de los resultados de su procesamiento. Dichos resultados brotan de los diversos

módulos como burbujas en una olla de agua hirviendo. Una tras otra, las burbujas —cada una de ellas, el resultado final del procesamiento de un módulo o un grupo de módulos— emergen y explotan durante un momento, para ser reemplazadas de inmediato por otras en un constante movimiento dinámico. Esos estallidos individuales de procesamiento desfilan uno tras otro perfectamente hilvanados por el tiempo (téngase en cuenta que esta metáfora se limita a las burbujas que brotan a una velocidad de doce fotogramas por segundo o más; piense en un folioscopio, uno de esos libros de dibujos animados en los que, cuanto más deprisa se pasan las páginas, más continuos parecen los movimientos de los personajes).

Sir Charles Sherrington, el decano de la neurociencia moderna, pensaba en algo similar cuando observó:

¿Hasta qué punto la mente es una colección de mentes perceptuales cuasi independientes integradas psíquicamente en gran medida por la concurrencia temporal de la experiencia? Sus reservas independientes de cerebro subperceptual y perceptual, si podemos llamarlo así, podrían explicar la levedad del deterioro mental derivado de algunas lesiones cerebrales. [...] La simple contemporaneidad puede ser una gran aglutinadora.³

Es difícil asimilar la idea de que cada burbuja tiene su propia capacidad de suscitar esa sensación de ser consciente, ya que choca con nuestra propia intuición respecto a la naturaleza holística de nuestra conciencia personal. ¿Qué se nos escapa, a nosotros y a nuestra intuición? Se nos escapa la parte de la ilusión, esa parte que con tanta facilidad suele escapárenos a los humanos (con nuestro potente mecanismo de inferencia del hemisferio izquierdo). No es que se nos escape exactamente la ilusión, sino el hecho de que la aparente continuidad del flujo de nuestra conciencia es una ilusión en sí; en realidad esta está integrada por burbujas cognitivas conectadas con burbujas «sensoriales» subcorticales, que nuestro cerebro hilvana a lo largo del tiempo.

El telón de fondo de las burbujas

Hay una observación clásica que parece cumplirse en toda la biología. Dicha observación se refiere a la cuestión de si los organismos aprenden y reciben instrucciones del entorno o si las reacciones que los organismos manifiestan ante los estímulos del entorno son gestionadas por sistemas previamente integrados en ellos. El debate «selección *versus* instrucción» se libra encarnizadamente desde hace años, y se ha convertido en foco de atención en el ámbito de la inmunología. En pocas palabras, cuando algo extraño penetra en el cuerpo y se produce una respuesta inmunitaria, ¿se forman los anticuerpos en ese momento en torno al cuerpo extraño y luego se multiplican (instrucción)?, ¿o bien el anticuerpo ya existe previamente y el tiempo de la respuesta inmunitaria es el tiempo que se necesita para identificar a ese anticuerpo preexistente (selección) y ponerlo en acción? En el siglo pasado, la biología descubrió que lo que ocurre realmente se corresponde con el segundo supuesto, un hallazgo que ilustra el hecho de que el paquete que constituye el equipamiento estándar de nuestro cuerpo y cerebro incluye un montón de cosas.

En 1967, el inmunólogo danés Niels Jerne postuló algo bastante sorprendente en aquel momento: lo que vale para el sistema inmunitario probablemente valga también para el cerebro. Así, sugirió que, durante lo que podríamos calificar como una situación de «aprendizaje», el entorno selecciona y hace uso de circuitos cerebrales preexistentes en el cerebro.⁴ Según esta potente visión naturalista, el aprendizaje es simplemente el tiempo que necesita el cerebro para revisar su infinidad de circuitos hasta encontrar el más apropiado para el reto en cuestión.

Aunque este debate crucial sigue vivo, nadie dudaría de que existen circuitos neuronales para funciones específicas que vienen como «equipamiento estándar» de nuestro cerebro. Así, por ejemplo, se ha demostrado que los bebés de solo seis meses ya poseen la capacidad de hacer inferencias causales.⁵ Y los circuitos subcorticales hacen su procesamiento evidente desde el momento en que el recién nacido llora para pedir su primera

comida. Siguiendo con la metáfora de la burbuja, diríamos que las burbujas son el resultado final del procesamiento de circuitos que están constantemente operativos para abordar y gestionar los interminables desafíos que plantea el entorno. Esos procesos son tanto corticales como subcorticales, pero antes de llegar a las burbujas subcorticales examinemos otro experimento revelador.

Murciélagos en el campanario

Para bien o para mal, la pregunta tristemente célebre de Thomas Nagel «¿Cómo es ser un murciélago?»⁶ ha servido de acicate a los estudiosos con mentalidad filosófica durante cuarenta años. En realidad la pregunta debería ser: «¿Qué tipo de burbujas tiene un murciélago?»; es decir, ¿cuáles son los contenidos de la conciencia de un murciélago? Probablemente nunca lleguemos a experimentar en plenitud la conciencia propia de un murciélago, pero sí podemos observar el contenido de un hemisferio solitario del cerebro humano. El cerebro está lleno de burbujas y, cuando se divide en dos, cada mitad se queda con su propio conjunto de burbujas en constante agitación. Dado que hoy sabemos que cada uno de los dos hemisferios tiene algunas burbujas únicas, ¿sería posible que cada hemisferio albergara un tipo distinto de experiencia consciente global? Para hacerse una idea al respecto, considere qué burbujas no tiene. Por ejemplo, yo puedo decir que no tengo la burbuja de la matemática abstracta y, en consecuencia, puedo afirmar también que me siento frustrado cuando empiezan a aparecer ecuaciones en las conferencias. Aunque desearía poder hacerlo, no puedo explicar cómo es entender las matemáticas más abstractas, ¡pero apuesto a que debe de ser genial!

Rebecca Saxe, del MIT, ha descubierto en algunos intrigantes estudios que existe un hardware cerebral especial en el hemisferio derecho del cerebro humano que parece estar especializado en determinar cuáles podrían ser las intenciones de otra persona.⁷ Cuando interactuamos con otros, estamos constantemente haciendo evaluaciones de manera refleja sobre su estado

mental y su intención en todas sus acciones. Es un proceso casi automático. Parece que los niños con autismo carecen en gran medida de esta capacidad y, como resultado, las interacciones sociales les resultan dificultosas. Como ya he mencionado antes, en el lenguaje psicológico formal esto se conoce como tener una «teoría de la mente». Utilizando modernas técnicas de captación de imágenes cerebrales, Saxe descubrió un área del hemisferio derecho responsable de esa capacidad. Como el lector habrá adivinado, esta observación plantea una nueva cuestión. El hallazgo de Saxe parece sugerir que quizá el hemisferio izquierdo de los pacientes con cerebro escindido no tenga acceso al módulo que añade la teoría de la mente a nuestra cognición. ¿Cómo sería un hemisferio izquierdo que no tuviera acceso a esa capacidad?

Michael Miller, antiguo alumno y ahora colega mío, y el distinguido filósofo Walter Sinnott-Armstrong unieron fuerzas para examinar las implicaciones del hallazgo de Saxe en los pacientes con cerebro escindido:⁸ querían determinar si un hemisferio separado podía evaluar las cuestiones morales de manera diferente al otro. Como hemos dicho, el trabajo de Saxe sugiere que en una persona con el cerebro escindido un hemisferio (el derecho) dispondría del módulo que evalúa la mente y las intenciones de los demás, mientras que el otro (el izquierdo) carecería de él. Tras la separación, ¿actuaría el hemisferio izquierdo de manera distinta al no contar ya con la colaboración de un módulo que evaluara los estados mentales y las intenciones del prójimo?

A los filósofos morales les gusta abordar los dilemas morales considerando que son de naturaleza o bien deontológica o bien utilitarista. Dicho en un lenguaje sencillo, la cuestión se resume en esta pregunta: «¿Resolvemos el dilema considerando lo que es intrínsecamente correcto (cuál es nuestro deber moral) o la solución reside en maximizar el bien colectivo?». Hay muchas formas de expresar esta dicotomía y muchas maneras de revelar si alguien es más deontológico o más utilitarista en su pensamiento. En una serie de pruebas ingeniosamente diseñadas, a los pacientes se les contaron historias cuyo

protagonista hacía algo malo, pero cuyo resultado, sin embargo, no causaba daño alguno: si una secretaria quiere cargarse a su jefe y se propone echarle veneno en el café, pero, sin que ella lo sepa, en realidad lo que le echa es azúcar, de modo que él se lo bebe y sigue estando perfectamente, ¿era eso permisible? En otros casos la historia era sobre una persona que hacía algo que parecía inocente, pero que resultaba fatal para otra: si una secretaria cree que está echando azúcar en el café de su jefe, pero en realidad es un veneno que ha olvidado accidentalmente un químico, y su jefe se lo bebe y muere, ¿era esa una acción permisible? El paciente, tras escuchar las historias completas, simplemente tenía que juzgar si la acción de aquella persona era «permisible» o «prohibida».

Huelga decir que la mayoría de la gente juzga el ejemplo en el que hay mala intención como prohibido, independientemente del resultado. En ese sentido, la mayoría de las personas son deontológicas. Y casi todas juzgan la acción de alguien que no tenía intenciones maliciosas como permisible (aunque no siempre), a pesar de que en ocasiones dicha acción pueda terminar en tragedia. Los pacientes con cerebro escindido, en cambio, actuaban de una forma peculiar. Al parecer, en un primer momento el hemisferio izquierdo —el «parlante»— ofrecía una respuesta utilitaria en todos los escenarios. En consecuencia, si una acción tenía mala intención pero no causaba ningún daño, se juzgaba como «permisible»; y si una acción no tenía mala intención pero causaba un daño, se juzgaba como «prohibida». Dada la nitidez de las historias utilizadas, era un resultado discordante. ¿Qué ocurría? El hemisferio izquierdo desconectado no podía tener en cuenta la intención de la persona en las historias, y actuaba como si no tuviera una teoría de la mente.

En segundo lugar, los pacientes solían dar explicaciones espontáneas acerca de por qué habían elegido la respuesta utilitarista por encima de la opción deontológica. Parecía que «sentían» que sus juicios no eran exactamente satisfactorios, y a menudo racionalizaban su opinión sin que nadie se lo pidiera. Recuerde que el hemisferio izquierdo tiene su «intérprete», el

módulo que trata de explicar tanto el comportamiento que observa que emana del cuerpo como las emociones que siente. Asimismo, hay que tener en cuenta que la reacción emocional a algo experimentado por un lado del cerebro la sienten los dos: si la emoción es el resultado de la experiencia del hemisferio derecho, el izquierdo no tendrá información acerca de por qué siente la emoción que está sintiendo, pero intentará explicarla de todos modos. Por lo tanto, cuando el hemisferio derecho oía la respuesta del izquierdo (aun con capacidades lingüísticas limitadas, todavía queda algo de comprensión en el primero), se quedaba tan sorprendido como nosotros, lo cual suscitaba una reacción emocional que no coincidía con lo que el hemisferio izquierdo consideraba una respuesta razonable. Tras sentar las bases de un importante conflicto como el que se planteaba, no resulta sorprendente que el módulo especial del hemisferio izquierdo (el módulo «intérprete») que siempre se apresta a explicar los comportamientos producidos por el derecho —ahora desconectado y silencioso— saltara a la palestra para intentar dar cuenta de lo que estaba ocurriendo. Por ejemplo, en uno de los escenarios planteados, una camarera servía semillas de sésamo a un cliente creyendo falsamente que le causarían una reacción alérgica dañina. El paciente J. W. juzgó la acción de la camarera «permisible»; pero al cabo de un momento ofreció una explicación espontánea: «Las semillas de sésamo son muy pequeñas. No hacen daño a nadie».

En mi metáfora, una burbuja es el resultado final del procesamiento de un módulo o un grupo de módulos en una arquitectura estratificada. En los pacientes con cerebro escindido, el módulo especial que evalúa la intención de los demás está desconectado y aislado del hemisferio izquierdo, que es el que está dotado del habla. En consecuencia, la burbuja del resultado de su procesamiento no emerge ya sea para contribuir al proceso de toma de decisiones del izquierdo o para luchar por asumir el control de dicho proceso. No puede formar parte de ese proceso de «burbujeo» por no estar ubicado físicamente en el hemisferio izquierdo, entre las burbujas que tienen acceso al

lenguaje y al habla; de modo que el conocimiento de la intención del otro desaparece de forma misteriosa. Ahora bien, las burbujas del procesamiento emocional del mesencéfalo llegan a ambos hemisferios. Solo cuando el hemisferio derecho oye la respuesta del izquierdo, que se traduce en una emoción que experimentan ambos hemisferios, el izquierdo identifica un desajuste, lo que desencadena el proceso de justificación. El hemisferio izquierdo también tiene toda una vida de recuerdos almacenados en relación con las normas morales de la cultura en la que ha crecido, y puede utilizarlos para justificarse.

Estamos descubriendo, pues, que las sutilezas de nuestra vida psíquica están gestionadas por módulos específicos de nuestro cerebro. Repitémoslo una vez más: el hemisferio izquierdo, que se beneficia de poseer los módulos que permiten el pensamiento abstracto, la codificación verbal y muchas otras cosas, no dispone, en cambio, del módulo necesario para tener en cuenta las intenciones del prójimo. Sin embargo, está dotado de una potente capacidad de inferencia. De ese modo, si el resultado es bueno, infiere que los medios eran correctos, y, por lo tanto, si el resultado es bueno la acción es permisible; si, por el contrario, el resultado último es malo, la acción no es permisible. Lo que es mejor para la mayoría es correcto. El aspecto más asombroso y casi surrealista de estos hallazgos es la posibilidad de que, si falta el módulo apropiado que nos permite pensar en los demás, al parecer resulta imposible aprenderlo.

¿Cómo es ser un hemisferio derecho?

¿Qué pasaría si de repente perdiéramos el hemisferio izquierdo? ¿Cuál sería nuestra experiencia consciente? Recuérdese que aparentemente no notaríamos que hubiera cambiado nada, puesto que, aunque parezca mentira, no echaríamos de menos los módulos del hemisferio izquierdo. Pero el hecho es que, habiendo perdido el centro del habla, las habilidades de comunicación y

comprensión se reducirían drásticamente. El hemisferio derecho solo sería capaz de gestionar una comprensión y un vocabulario limitados. No obstante, uno de los factores que más alterarían nuestra experiencia sería la pérdida de la capacidad de hacer inferencias. Utilizamos constantemente esa capacidad del hemisferio izquierdo, y la falta de ella nos sumiría en una experiencia del mundo muy distinta. Aunque sabríamos que los demás tienen intenciones, creencias y deseos, y podríamos intentar adivinar cuáles podrían ser, no podríamos inferir causa y efecto. No podríamos inferir «por qué» alguien está enfadado o cree lo que cree. Muchos de nuestros encuentros sociales probablemente resultarían en malentendidos y frustración por ambas partes. Pero la pérdida de la capacidad de inferencia no se limita al mundo social. No seríamos capaces de inferir ninguna causalidad en absoluto. No solo no podríamos deducir que nuestro vecino está enfadado porque nos dejamos la puerta abierta y su perro se escapó; es que ni siquiera seríamos capaces de inferir que el perro se escapó porque habíamos dejado la puerta abierta. No entenderíamos que el coche no arranca porque nos dejamos la radio encendida.

Aunque se nos darían muy bien las relaciones espaciales, no comprenderíamos las causas y los efectos descritos por la física. No podríamos inferir ninguna fuerza causal no observada, ya fuera de índole gravitacional o espiritual. Por ejemplo, no deduciríamos que una bola se movía porque otra que había chocado con ella le había transferido una fuerza; en cambio, gracias a nuestra incapacidad para hacer inferencias, nos iría mejor en las mesas de juego de los casinos: apostaríamos con la casa y no trataríamos de inferir ninguna relación causal entre ganar y perder más allá del azar. Nada de corbata o calcetines de la suerte, ni de inclinar la cabeza de una determinada manera u otras cosas por el estilo. No inventaríamos historias descabelladas acerca de por qué hicimos tal cosa o sentimos tal otra, no porque no tuviéramos la capacidad de lenguaje, sino, una vez más, porque no podríamos inferir causa y efecto. No seríamos hipócritas ni pretenderíamos

racionalizar nuestras acciones. Tampoco podríamos deducir lo esencial de las cosas, y lo tomaríamos todo en sentido literal. No entenderíamos las metáforas ni las ideas abstractas. Sin inferencia, estaríamos libres de prejuicios, pero la imposibilidad de inferir causa y efecto nos dificultaría el aprendizaje. Las burbujas de procesamiento que emergen en cada uno de nuestros hemisferios por separado determinan cuál será el contenido de la experiencia consciente de ese hemisferio.

Sentimientos

Todos hemos experimentado el anhelo de volver a un lugar favorito del mundo, un lugar del que una vez disfrutamos y que recordamos con tanto cariño que sentimos que tenemos que volver allí y recuperar el momento. Sin embargo, cuando volvemos a visitarlo por segunda vez la experiencia no es exactamente igual. Los sentimientos son distintos. No es que sean mejores ni peores. Simplemente son distintos.

Hace poco mi esposa y yo regresamos a la población italiana de Ravello, un lugar que en el pasado nos había parecido mágico. Todavía conservaba su belleza natural. Todavía tenía su historia, su cultura y, lo mejor de todo, su gente; sin embargo, nos causó una impresión distinta. El lugar no parecía coincidir con nuestros sentimientos, enraizados en nuestra experiencia previa. ¿Acaso nos traicionaba el recuerdo, o era que teníamos un «sentimiento» distinto sobre la vida en general y ese sentimiento distinto teñía nuestra experiencia actual?

Nuestra forma habitual de concebir las experiencias pasadas es que cada una de ellas tiene un sentimiento asociado que viene determinado por el acontecimiento en cuestión. El sentimiento real en sí está ligado a la experiencia real, y esperamos que, reproduciendo la experiencia, podamos recuperar ese mismo sentimiento. Estar de vacaciones en el sitio *x* nos causó una gran felicidad y emoción; si regresamos, volveremos a sentir esa misma

emoción y felicidad. Esta podría ser la razón por la que la gente compra multipropiedades. La primera visita es genial, pero la siguiente... ¿y la siguiente? Repetimos el acontecimiento y esperamos que el sentimiento original reaparezca de forma automática. Pero no es esta la forma correcta de concebirlo.

Ese momento mágico del que disfrutamos en el pasado actualmente está almacenado como información sobre dicho acontecimiento pretérito. Se coloca en la memoria de formas que todavía no hemos llegado a entender, pero es información simbólica, fría y formal, del mismo modo que el ADN es información simbólica; y, al igual que el ADN, tiene una estructura física. Esa estructura informativa puede contener el hecho de que el acontecimiento estuvo asociado a una serie de sentimientos positivos; pero los sentimientos en sí no se almacenan, solo la información sobre ellos. Cuando emerge una burbuja de memoria, también lo hacen las burbujas que expresan nuestro sentimiento «actual» al respecto. Los sentimientos provienen de otro sistema, con sus propias burbujas de procesamiento, independientes del sistema de la memoria. En resumen, lo que estoy sugiriendo es que el desacoplamiento de la dimensión emocional revela el hecho de que hay diferentes sistemas que convergen en el tiempo para producir un sentimiento relativo a un recuerdo. Por hacer una analogía, es como la banda sonora musical de una escena de una película dramática. La imagen y la música están separadas, pero cuando se unen la banda sonora aporta emoción. A medida que las burbujas se suceden rápidamente unas a otras, se crea la ilusión de albergar sentimientos sobre el acontecimiento recordado.

Así pues, tenemos burbujas de memoria y burbujas de sentimiento. Cuando pensamos en el momento pasado, nuestros sentimientos al respecto no son realmente los sentimientos de aquel momento, sino que son sentimientos actuales que proyectamos sobre dicho momento pretérito. En general, son positivos, pero los sentimientos en sí no están fijados a los recuerdos pasados reales, lo cual resulta más evidente cuando el sentimiento actual es el opuesto

al original. Por ejemplo, pongamos por caso que recordamos el momento en el que nos dieron los resultados de un examen que suspendimos en la universidad. Sin duda, nuestra memoria nos dice que en aquel momento nos sentimos mal; pero supongamos que gracias a aquel fracaso acudimos al despacho del profesor a pedirle ayuda, lo que de hecho nos llevó a meternos más de lleno en el tema, lo cual, a su vez, permitió que termináramos haciendo carrera en aquella disciplina. Cuando pensamos retrospectivamente en el hecho de aquel suspenso, hoy, que sabemos a dónde nos condujo, solo podemos pensar en ello en términos positivos. Recordamos intelectualmente que entonces nos sentimos mal, pero no podemos reencarnar ese sentimiento. Posiblemente el sentimiento más resistente a los cambios de perspectiva sea el de vergüenza. Puede que todavía nos sonrojemos pensando en alguna transgresión social pasada, pero a veces no podemos menos que sonreír y menear la cabeza al recordar lo que nos avergonzó de jóvenes (o de adolescentes); por ejemplo, cuando en el autocine nos escondíamos en el suelo de la parte trasera del coche familiar para que nadie viera que íbamos con nuestros padres. ¿De verdad hice eso? Recordarlo ya no nos sonroja... o quizá sí, aunque por una razón completamente distinta.

Los módulos que producen las «sensaciones crudas» son diferentes de los que generan los pensamientos, los recuerdos, las decisiones, etc. Lo que sentimos en un momento del tiempo durante un acontecimiento concreto se convierte en un aspecto del recuerdo de dicho acontecimiento, en una dimensión, una información que podemos etiquetar y codificar neuronalmente e incluir en nuestra memoria.

El sentimiento real, en cambio, proviene de otro «contratista» independiente del cerebro. En una nueva visita a un lugar, este puede tener una configuración neurológica diferente a la de la visita anterior. En ambas ocasiones, es esa configuración neurológica la que motiva nuestro sentimiento respecto a la visita. La primera visita conlleva el reto de lo desconocido, la curiosidad, la aventura, a menudo la juventud y, a veces, un subidón de

adrenalina. En la nueva visita, uno regresa con más años, con más experiencia vital en su haber, a un territorio que ya le resulta familiar; no es tan estimulante, no nos espolea tanto la curiosidad, sabemos más o menos lo que podemos esperar y nos sube menos la adrenalina. Albergamos un sentimiento distinto; no necesariamente peor o mejor, solo distinto. Eso no solo vale para cuando regresamos a lugares físicos, sino también para cuando intentamos recuperar cualquier experiencia pasada. Neil Young lo describe muy bien: «Todavía intento ser así, pero, ya sabes, ya no tengo veintiuno ni veintidós. [...] No estoy seguro de poder recrear ese sentimiento, [ya que] tiene que ver con la edad que tenía entonces, lo que sucedía en el mundo, lo que yo acababa de hacer, lo que quería hacer a continuación, con quién vivía, quiénes eran mis amigos, qué tiempo hacía». ⁹

Sentio ergo sum

Hace algunos años, Steven Pinker comentaba: «Hay algo en el tema de la conciencia que hace que las personas, como la Reina Blanca de *A través del espejo*, crean seis cosas imposibles antes del desayuno. ¿Es posible que la mayoría de los animales sean realmente *inconscientes*: sonámbulos, zombis, autómatas, pasmados? ¿No tiene un perro sentimientos, afectos, pasiones? Si los pinchas, ¿no sienten dolor?». ¹⁰ Jaak Panksepp coincidía con Pinker en que negar eso es creer en lo imposible, y culpaba de ello a Descartes, considerando que estaba totalmente fuera de lugar que negara la conciencia a los animales. Es más: creía que Descartes nos habría ahorrado muchos problemas si, cuando se preguntó: «¿Qué es ese “yo” que “yo” conozco?», se hubiera limitado a responder: «Siento, luego existo», y hubiera dejado la cognición fuera de la ecuación de la experiencia subjetiva. Panksepp también habría coincidido con Pattee en que, desde Descartes, casi todo el mundo ha estado buscando demasiado arriba en el árbol evolutivo para averiguar cómo

se las arreglan los sistemas neuronales para producir la experiencia afectiva subjetiva. ¹¹

Panksepp postulaba que la experiencia afectiva subjetiva surgió cuando el sistema de las emociones —evolutivamente antiguo— se interconectó con una versión primitiva de «mapa corporal» neuronal que delimitaba el «yo» con respecto al mundo externo. ¹² La formación de ese mapa corporal solo requeriría que las sensaciones de dentro y fuera del cuerpo se «pegaran» en las «hojas» de las correspondientes neuronas del cerebro. Asimismo, sostenía que los dos elementos necesarios para la experiencia subjetiva eran la información sobre los estados internos y externos del agente (registrada simbólicamente) y la construcción de una simulación neuronal integrada del agente en el espacio: un modelo rápido e imperfecto, construido a partir de las estimulaciones neuronales. Información y construcción: la misma complementariedad que veíamos en el ADN. La cognición o el conocimiento superiores de que tenemos un «yo» (esto es, la «autoconciencia») no forma parte de la receta original. Para moverse por el entorno de forma segura y eficiente, comer cuando se tiene hambre, etc., uno no necesita saber que es consciente de sí mismo, pero sí necesita conocer los límites de su cuerpo en relación con el espacio que lo rodea. Si no los conociera, estaría constantemente tropezando con todo y calibrando mal todo tipo de cosas, desde saber si cabe o no en una acogedora guarida hasta determinar si puede saltar de una roca a otra. Y también necesita tener una motivación para actuar de formas que favorezcan la supervivencia y la reproducción. Es decir, que no necesitamos que emerjan las burbujas de una corteza cerebral altamente desarrollada para ser conscientes de una experiencia subjetiva. Descartes solo necesitaba las señales de la subcorteza para sentir que era un «yo»; no necesitaba para nada pensar en ello.

De hecho, hasta un insecto necesita tener información sobre su cuerpo en el espacio para poder moverse de manera segura y eficaz. En ese aspecto me he visto superado por muchas moscas, empuñando mi matamoscas sin efecto

alguno. El biólogo Andrew Barron y el filósofo de la neurociencia Colin Klein, ambos de la Universidad Macquarie de Sídney, empezaron a indagar en el mundo del cerebro de los insectos, y descubrieron que las estructuras conocidas como «complejo central» realizaban funciones análogas a las de algunas partes del mesencéfalo de los vertebrados, generando un «modelo espacial unificado del estado y la posición del insecto en el entorno». ¹³ Es decir, que las funciones que ubican a un organismo en el espacio están presentes en los cerebros de las cucarachas y los grillos, las langostas y las mariposas, las moscas de la fruta y las abejas, como lo están en el mesencéfalo de los vertebrados. En suma: la tarea de ordenar la plétora de sistemas necesarios para realizar una acción es una característica biológica extendida en los organismos complejos. La tienen los gusanos y la tenemos nosotros. Barron y Klein, coincidiendo con Panksepp y Merker, concluyen: «Esta representación integrada y egocéntrica del mundo desde la perspectiva del animal basta para la experiencia subjetiva». También consideran que esa conciencia del cuerpo en el espacio es suficiente para los insectos que han estudiado, lo que sugiere que también ellos tienen una forma de experiencia subjetiva, que estaba presente en un ancestro común de los vertebrados e invertebrados en la Explosión Cámbrica, hace unos quinientos cincuenta millones de años.

No son los únicos que piensan así. Los neurobiólogos Nicholas Strausfeld, de la Universidad de Arizona, y Frank Hirth, del King's College de Londres, también se han aventurado en su propia rama del árbol evolutivo. En su investigación, realizaron un exhaustivo examen de las características anatómicas, de desarrollo, conductuales y genéticas de los ganglios basales de los vertebrados y las compararon con las del complejo central de los artrópodos (de los que los insectos son solo una rama). Encontraron tantas similitudes que no pudieron por menos que concluir que los circuitos del complejo central de los artrópodos y de los ganglios basales de los vertebrados compartían un ancestro común. ¹⁴ De hecho, ambos se derivan de

un programa genético evolutivamente preservado; es decir, que los circuitos esenciales para las opciones conductuales se originaron muy temprano en la evolución. El ancestro que los vertebrados compartimos con los artrópodos ya correteaba de un lado a otro con ese tipo de circuitos y, por lo tanto, ya disponía de burbujas emergentes de información procesada sobre su ubicación y sensaciones que guiaban sus acciones. Strausfeld y Hirth sugieren incluso que el equipamiento cerebral del ancestro común bastaba para sustentar la experiencia fenoménica. Aunque su punto de vista pueda resultar algo exagerado, lo cierto es que su trabajo nos alerta de hasta qué punto los mecanismos básicos que estamos descubriendo en los humanos pueden haberse preservado evolutivamente desde tiempos muy remotos. Esa es la belleza de la investigación evolutiva y comparada: toda una serie de aspectos de nuestra propia vida mental que presuponemos construcciones únicas del cerebro humano resulta que existen desde hace mucho tiempo por más que hayan sido especialmente cultivados por nuestro cerebro.

¿Burbujas metódicas o caóticas?

Nuestra experiencia consciente es un flujo constante e ininterrumpido de pensamientos y sensaciones. ¿Cómo puede ser así, con todas esas burbujas luchando para tener prioridad sobre las demás? ¿Emergen las burbujas de cualquier manera o son el producto de un sistema de control dinámico? ¿Hay una capa de control que dé su aprobación a algunas burbujas al tiempo que acalla a otras?

Una de las formas en que se controla el procesamiento de un módulo es mediante el estímulo de entrada que recibe. Pongamos por caso que damos un mordisco a una trufa de chocolate hecha de chocolate amargo. Las células gustativas que detectan el sabor dulce no activarán ningún nervio aferente —es decir, procedente de la periferia del cerebro—, de manera que tampoco se activará ninguno de los módulos encargados de procesar dichas sensaciones,

por lo que no experimentaremos ningún sabor dulce ni se procesará ninguna información relacionada con dicho sabor. En cambio, las células encargadas de detectar el sabor amargo sí se activarán, y notaremos la boca inundada de ese sabor; se está procesando la información relativa al sabor amargo. Si a continuación sustituimos la pieza de chocolate amargo por una de aspecto idéntico pero de chocolate con leche, el módulo que procesa el sabor dulce también entrará en funcionamiento, enviando una burbuja de dulzor a nuestra conciencia a toda máquina y superando con creces al sabor amargo. Es como si ahora el sabor amargo fuera solo un recuerdo lejano y el sabor dulce se haya adueñado del momento, al menos hasta que emerja la siguiente burbuja. Aquí no se requiere para nada la cognición. De alguna manera, esa señal dulce ha ganado la competición de burbujas. ¿Ha recibido alguna ayuda extra?

Se han detectado algunas formas de potenciación selectiva de ciertas señales en diversas especies de animales —desde los cangrejos ¹⁵ hasta las aves ¹⁶ y los primates ¹⁷—, lo cual sugiere que se trata de una habilidad compartida con nuestro último ancestro común, que corría por el mundo hace unos quinientos cincuenta millones de años. La primera manifestación de esta capacidad de «control de datos» fue una forma rudimentaria de atención, un proceso que ayudó a gestionar la arremetida de información sensorial que desbordaba a un grupo de células exploradoras. En una fase evolutiva temprana apareció un proceso de potenciación de señales para ayudar a los organismos a determinar cuáles de entre todos los estímulos que los bombardeaban podrían ser más relevantes para la supervivencia (era mejor dar prioridad a la información sobre una amenaza, comida o pareja que a muchas otras cosas), y ese proceso se ha preservado en todas las formas de vida derivadas de ese organismo primitivo.

Steven Wiederman y David O'Carroll, de la Universidad de Adelaida, descubrieron que en el cerebro de la actual libélula hay una única neurona visual que selecciona un determinado objetivo que parezca una posible presa y lo sigue, ignorando por completo a otros. ¹⁸ Esto resulta interesante no solo

porque demuestra que las libélulas poseen una forma de selección competitiva, necesaria para la atención visual, sino también porque ese proceso de atención selectiva lo realiza una sola célula. En los vertebrados, la potenciación selectiva de señales evolucionó hasta desembocar en lo que hoy denominamos «atención», un sofisticado mecanismo que controla el flujo de datos entrantes y, a través de este, nuestra mente. Gracias a eso, por ejemplo, aunque podamos estar absortos viendo una película emocionante, si suena una alarma contra incendios nuestro sistema de «control de datos» potenciará de inmediato ese chirriante estímulo sonoro, apartando nuestra mente de la película y aprestándonos para entrar en acción.

Sin embargo, aunque la potenciación selectiva de señales ejerce un cierto control sobre nuestra mente, esta última también controla en alguna medida nuestra atención. Lo que quiero decir es que la atención tiene dos componentes: uno que actúa en sentido ascendente y otro que lo hace en sentido descendente. Si acudimos a una cita a ciegas y sabemos que ella llevará una flor roja en el pelo, nuestros ojos se moverán de un lado a otro mirando peinados y poca cosa más. Nuestra atención está funcionando en sentido descendente, esto es, guiada por nuestra intención de encontrar a la desconocida. La atención ascendente no nos lleva mucho más lejos en el juego de la supervivencia, de modo que los organismos que desarrollaron una atención descendente tenían una clara ventaja, y esta pasó a convertirse en un elemento necesario. Esa capacidad de atención descendente está altamente desarrollada tanto en las aves como en los mamíferos, cuyo ancestro común circulaba hace unos trescientos cincuenta millones de años, de modo que se trata de una habilidad al menos tan antigua como la atención ascendente, aunque más novedosa desde un punto de vista evolutivo. Fue una extensión añadida: una nueva capa.

También aquí Pattee formula una provechosa hipótesis acerca de cómo pudo evolucionar esa capa. El fracaso de una capa es la fuerza o condición básica para que surja otra nueva:

Cuando un sistema no logra tener una representación o una descripción para manejar una situación determinada, deja un vacío de poder, por así decirlo, o un vacío de decisión. Yo lo calificaría como una especie de inestabilidad en cuanto que se necesita tomar una decisión pero no existe un procedimiento decisorio. Lo que se tiene entonces es ambigüedad, y las pequeñas causas pueden tener grandes efectos. Se trata, de hecho, de una crisis del sistema, y entonces puede surgir un nuevo tipo de comportamiento.¹⁹

Cuando los sistemas se hicieron más complejos surgió la necesidad de disponer de algún tipo de capa de control que gestionara la plétora de estímulos independientes y el comportamiento resultante. Aunque la potenciación de estímulos es genial, hacía falta una capa de control que orquestara de alguna forma el bullicio de todos los módulos.

Una famosa disfunción de la capa de control

Como ya hemos mencionado anteriormente, las personas que sufren lesiones en el lóbulo parietal derecho pueden experimentar una heminegligencia de la parte izquierda del campo visual incluso en sus imaginaciones y recuerdos. El primero que realizó esta conocida observación fue el brillante neuropsicólogo italiano Edoardo Bisiach.²⁰ En pocas palabras, lo que hizo fue pedir a sus pacientes que describieran de memoria la Piazza del Duomo de Milán desde dos perspectivas distintas. Se trata de un lugar muy hermoso, con edificios peculiares alineados a cada lado de la plaza y la gran catedral al fondo. Todo el que vive en Milán puede describirla con facilidad.

Cuando se pidió a los pacientes que describieran el Duomo como si estuvieran delante mirándolo de frente, lo hicieron sin ningún problema; pero solo describieron los edificios que se alineaban en el lado derecho (norte) de la plaza, obviando cualquier descripción de los del lado izquierdo (sur). A continuación Bisiach les pidió que se imaginaran que habían dado un giro de ciento ochenta grados y que describieran la plaza como si estuvieran en las escaleras de la entrada del Duomo. Los mismos pacientes describieron sin problemas los edificios del lado derecho de la plaza (sur) mirando desde el

Duomo, ignorando por completo los del lado izquierdo (norte), pese a que acababan de describirlos cuando se imaginaban situados en la dirección opuesta.

Este espectacular ejemplo clínico revela la existencia de dos conjuntos de módulos completamente diferentes. Resulta obvio que los módulos encargados de generar la imagen mental permanecen intactos y que toda la información está presente, pero esos módulos están controlados por un módulo adicional que evalúa desde qué lado del espacio se describirá la imagen, y es ese módulo el que presenta una disfunción. Años después de este experimento, la neuróloga Denise Barbut y yo tuvimos ocasión de examinar el caso de un paciente del Hospital de Nueva York con daños similares en el lóbulo parietal derecho, obteniendo los mismos resultados que Bisiach.²¹

Una conciencia enriquecida por la evolución

Cuando se trata de la conciencia, parece que hayamos olvidado el hecho de que nuestro cerebro evolucionó acrecentando su complejidad. En el transcurso de la evolución se han ido añadiendo módulos y capas con el tiempo para resolver una perturbación tras otra, cambiando e incrementando el contenido de nuestra experiencia consciente a lo largo del camino. Cada capa tiene sus propias reglas independientes de procesamiento y transmite su trabajo, su burbuja de procesamiento, a la siguiente. Mientras que el procesamiento dentro de un módulo, de una capa a otra, puede ser secuencial, los distintos módulos trabajan en paralelo, y de cada uno de ellos emergen sus propias burbujas hacia su destino final.

En la analogía de las burbujas, los resultados del procesamiento de varios módulos irrumpen en nuestra conciencia de un instante a otro. Lo más probable es que cada burbuja se vea potenciada al primer plano por una capa de control dotada de un protocolo formado por reglas arbitrarias, reglas que la evolución ha seleccionado porque proporcionaban al contenido de la conciencia la

información más fiable y adecuada para la situación que se afrontaba. Pero basta que aparezca una regla mejor y más fiable, y el protocolo puede cambiar. Por ejemplo, consideremos el caso de una burbuja de creencia que aflora a la conciencia. Supongamos que creemos que las grasas saturadas no son saludables y que ingerirlas nos está haciendo ganar peso; creemos que deberíamos reemplazar esas calorías con carbohidratos, y lo creemos porque es lo que nos han dicho los funcionarios de salud pública, los nutricionistas y nuestro médico. Así, cuando vamos a comprar al súper, emergen burbujas que nos guían para mantenernos alejados de las grasas saturadas. Pero más tarde observamos un hecho: nuestra experiencia no respalda la aserción de los funcionarios y demás. Cuanta más grasa eliminamos de nuestra dieta reemplazándola por carbohidratos, más engordamos. A continuación leemos un libro cuyo autor revisa y evalúa todas las investigaciones relacionadas con aquella aserción, y descubre que no solo la mayoría de dichas investigaciones no cumplen con los estándares de las buenas prácticas científicas, sino que además las pocas que sí lo hacen no respaldan lo que se nos ha dicho; es más, sugieren que lo que sucede es justo lo contrario.²² Al final nos convencemos. Empezamos a comprar e ingerir mantequilla y crema de leche. Perdemos peso. En el súper, las burbujas de la crema toman el control. Ha aparecido una regla mejor y más fiable, y el protocolo ha cambiado. Además, el café sabe mejor.

Es importante destacar que las brechas, los módulos y las capas pueden ayudarnos a comprender el comportamiento que se observa en las personas que sufren lesiones cerebrales. Si perdemos parte del tejido neuronal que procesa determinados aspectos concretos de la información, esa información dejará de formar parte de nuestro repertorio de burbujas y ya no aportará contenido a nuestra experiencia consciente. Lo mismo vale cuando el hemisferio derecho está separado del izquierdo: ninguno de los dos hemisferios cuenta ya con las burbujas que emergían del hemisferio opuesto para enriquecer su experiencia consciente, con lo que la experiencia consciente de cada uno de los dos hemisferios resulta empobrecida.

La conciencia es un instinto

La «magia» de un hombre es la ingeniería de otro.

ROBERT A. HEINLEIN

Cuando yo era joven, justo cuando iniciaba mis estudios de posgrado en el Instituto de Tecnología de California, trabé amistad con el filósofo político Willmoore Kendall, que tenía una personalidad discordante donde las hubiere (de hecho, era tan exasperante que la administración de Yale le había pagado una importante suma de dinero para que dimitiera). Kendall cuestionó todos los presupuestos de la institución sobre casi todo, y luego se mudó al oeste del país. El oeste no le era desconocido, ya que había nacido en la ciudad de Konawa, Oklahoma, hijo de un pastor de la Iglesia ciego. Finalmente, después de Yale, se estableció en una pequeña escuela jesuita de Dallas. Su apetito por la vida, que no tenía límites, era consecuencia de un ego firmemente arraigado. El día que dispararon a Kennedy en Dallas recibí una llamada suya. Me dijo: «Nunca antes había estado en un almuerzo donde hablara el presidente de Estados Unidos. Debería haber imaginado que pasaría algo».

Kendall no dejaba de estimular mi mente, ya que me veía como una víctima de lo que él consideraba el exacerbado pensamiento reduccionista moderno. Yo estaba convencido —y a día de hoy sigo estándolo— de que el mecanicismo físico puede explicarlo casi todo y de hecho lo hará. En general, cuando los filósofos empiezan a analizar el pensamiento fundamental, los científicos de laboratorio ponen los ojos en blanco. Kendall luchaba contra alguien que no estaba preparado para una lucha intensa y que de hecho apenas

era consciente de que hubiera un problema. Dado que yo le dejaba siempre mi apartamento cada vez que iba a pasar el día a Pasadena, él decidió que debía corresponderme y proporcionarme una educación más extensa. En ese contexto me recomendó que leyera el libro clásico de Michael Polanyi *Personal Knowledge*, que surgió a partir de las Conferencias Gifford que el autor pronunciara en 1951. Y lo hice. Desde entonces la obra ha permanecido siempre en los recovecos de mi mente, al igual que en los de mi estantería, arrastrada de un lado a otro del país en incontables ocasiones.

Polanyi —solía decirme Kendall— era un auténtico erudito. En 1916 escribió su tesis doctoral de química durante un permiso por enfermedad mientras servía como médico en el frente serbio. Aunque obtuvo una cátedra de química física en la Universidad de Manchester, la amplitud de sus intereses, que abarcaba materias como la economía, la política y la filosofía, llevó a la universidad a crear una cátedra de ciencias sociales expresamente para él. «¿Sabes? —me decía Kendall—, cada día responde a su correspondencia en doce idiomas distintos.» Polanyi era un gigante de su época y, aunque estaba establecido en Inglaterra, se desplazaba regularmente a la Universidad de Chicago como profesor visitante. La lectura de su libro me hizo ser consciente de la espinosa cuestión de que el conocimiento de las partes de una cosa no siempre te dice cómo podría ser el todo. Hay algo más en juego, falta algo; y creo que ese enfoque es el origen de la que llamaré «la escuela de pensamiento de Chicago», que planteaba su propia perspectiva con respecto a los procesos cerebrales.

El hecho de que «faltara algo» era un resultado directo de la metáfora de la máquina, que existía desde los tiempos de Descartes y cuyo anzuelo se habían tragado los biólogos hasta el fondo. Los científicos de Chicago se habían dado cuenta de que la tradicional analogía determinista clásica de la máquina para representar la vida en realidad funciona exactamente al revés: no es que los cerebros sean como las máquinas, sino que las máquinas son como los cerebros, aunque les falta algo. Polanyi precisaba que los humanos

evolucionaron mediante la selección natural, mientras que las máquinas están fabricadas por humanos; existen solo como producto de la materia viviente altamente evolucionada y representan el producto final de la evolución, no su comienzo.

La escuela de Chicago también lanzó la idea de que el origen de la vida dependía de dos modos complementarios de descripción y no solo de la descripción proporcionada por la física clásica que tan bien funciona para las máquinas. Pattee lo resumió así: «La vida en sí misma no podría existir si dependiera de tales descripciones clásicas o de que realizara sus propios procesos de registro interno de esa forma clásica». ¹ Esto venía a dar continuidad a Rosen, que ya había removido las aguas anteriormente al preguntarse: «¿Por qué no iba a ser posible que los “universales” de la física lo sean solo en una pequeña clase especial (aunque desmesuradamente prominente) de sistemas materiales, una clase de la que los organismos son demasiado “generales” para formar parte? ¿Y si la física es lo particular y la biología lo general, en lugar de ocurrir al revés?». ²

La batalla contra el reduccionismo puro se inició con Polanyi y el profesor de la Universidad de Chicago Nicolas Rashevsky, el padre de la biofísica matemática y la biología teórica, que no dejaba de ser un insólito soldado en aquella lucha. En un primer momento, Rashevsky había decidido abordar el problema de establecer la base material de los fenómenos biológicos básicos en general a raíz de la indignación que sintió cuando en una fiesta un biólogo le dijo que nadie sabía cómo se dividían las células, y que eso era algo que nadie podía saber porque era una cuestión de biología y, por lo tanto, estaba fuera del alcance de la física. Después de trabajar arduamente en el problema durante las décadas de 1930 y 1940, se sentía cada vez más incómodo. Como describe Rosen, que fue alumno suyo: «Se había hecho la pregunta básica “¿Qué es la vida?”, y la había abordado desde un punto de vista tácitamente tan reduccionista como cualquiera de los biólogos moleculares actuales. El problema era que, al tratar las funciones individuales de los organismos y

captar esos aspectos en modelos y formalismos independientes, de alguna manera había perdido de vista los propios organismos en sí, y ya no pudo recuperarlos». ³ De modo que llegó a la conclusión de que «no había ninguna colección de “descripciones” independientes [es decir, “modelos”] de organismos, por muy exhaustivos que fueran, que unidos pudieran captar el organismo en sí. [...] Si se pretendía alcanzar ese objetivo, se necesitaba un nuevo principio». ⁴ Rashevsky denominó «biología relacional» a la búsqueda del nuevo principio. En muchos aspectos, mi mentor Roger Sperry, que se formó en la Universidad de Chicago, también hizo suya aquella búsqueda. Y, como hemos visto, estas ideas también tuvieron una profunda influencia en Howard Pattee, a quien se atribuye el mérito de traer esta línea de razonamiento a la época actual.

El persistente mensaje de aquella temprana escuela de Chicago era este: a la hora de considerar un organismo, hay algo más que debe tenerse en cuenta. El pensamiento mecanicista está bien y nos enseña todo lo necesario sobre las partes, las capas de procesos automáticos que trabajan incansablemente en cualquier organismo para permitir su existencia. Pero hay «algo más», otro factor que es necesario entender, y ese algo no es un fantasma en el sistema: es el sistema, el organismo en sí, que puede modular las capas inferiores que lo producen. Es lo que responde a la pregunta: «¿Qué es la vida?».

Como argumenta Rosen, la ciencia siempre introduce un sustituto (un modelo) del objeto real que trata de estudiar. Con ese sustituto, los científicos pueden utilizar todos los métodos de la ciencia reduccionista y descubrir cómo funcionan las partes. El supuesto es que el sustituto puede reemplazar al objeto real. Pero, cuando vuelven al objeto real después de haber estado trabajando en el sustituto y tratan de aplicarle sus hallazgos, normalmente se quedan cortos. Por ejemplo, una cosa es estudiar el páncreas por sí solo en una placa bajo un microscopio o en un tubo de ensayo: eso puede enseñarnos cómo funciona a escala local; pero si no lo estudiamos conectado al cuerpo, no entenderemos su función real, ni cómo funciona en concierto con un sistema

distante (en este caso, una parte del intestino) que a la vez lo modula. El hecho de que las funciones del páncreas y del intestino están interrelacionadas no se descubrió hasta que los cirujanos empezaron a realizar operaciones de banda gástrica para reducir la obesidad y observaron que la diabetes desaparecía de la noche a la mañana. Para la neurociencia, el sustituto que reemplazó al cerebro era «una máquina», y al concebirlo como máquina pasaron por alto todo el concepto de complementariedad y lo que este nos aporta para comprender cómo funciona el cerebro.

Sperry lo expresó de manera distinta. Cuando sugirió que nuestras capacidades mentales eran entidades reales y formaban parte de la cadena causal de acontecimientos que desembocan en el comportamiento, los reduccionistas se subieron por las paredes, como ya hemos expuesto en el Capítulo 3. Sin embargo, él tampoco consideraba que los acontecimientos mentales, como los pensamientos, fueran acontecimientos no físicos o fantasmas en el sistema, sino que los concebía como el producto de las propiedades configurativas de los circuitos neuronales subyacentes. Dichos circuitos subyacentes tenían una estructura tanto física como simbólica, que controlaba lo que se estaba construyendo: el acontecimiento mental; son los símbolos físicos de Pattee controlando la construcción. En suma, Sperry creía que el organismo en sí desempeñaba un papel en su propio destino. Desde esta perspectiva, ni aun sabiendo todo lo que se puede saber sobre el estado actual de nuestro cerebro —sus condiciones iniciales— podríamos predecir cómo los estados mentales futuros pueden tener un efecto descendente en nuestro procesamiento ascendente. Esas condiciones iniciales no nos dirán qué, dónde y con quién cenaremos dentro de un año a partir del jueves. Saberlo todo sobre el estado del cerebro de un recién nacido no nos permitirá saber lo que hará ese niño un martes por la tarde cuarenta y cinco años después, tal como creen los más acérrimos deterministas. De hecho, ese determinismo extremo resulta casi tan necio como la creencia manifestada por el problema del gato de Schrödinger.

Mirando hacia el futuro

En nuestro esbozo de la historia del pensamiento y la investigación sobre el problema de la conciencia, hemos observado numerosos subterfugios. Solo a partir de Descartes y del nacimiento de la idea de que «el cerebro es una máquina que puede entenderse desmontándola» (la condición *sine qua non* del enfoque científico de cualquier cosa) arraigó con firmeza la ferviente devoción al reduccionismo que todavía sigue siendo la idea dominante en la neurociencia actual. Ya hemos dicho que la escuela de Chicago —como yo la llamo— ha echado el freno y ha señalado el camino hacia otra formulación, que tiene en cuenta la naturaleza evolutiva del organismo y el hecho de que las máquinas son subproductos del cerebro humano (mientras que los cerebros no son subproductos de las máquinas). Hay algo distinto en la materia viviente. Para decirlo sin rodeos, es el hecho de que esta no obedece solo las órdenes de las interacciones físicas clásicas, sino que tiene asimismo una arbitrariedad innata conferida por la información simbólica física, aunque arbitraria, que reside en el «lado bueno» del *Schnitt*.

Cuando se conocieron y establecieron los primeros resultados de la investigación en torno al cerebro escindido, la cuestión persistente pasó a ser: ¿qué nos enseña esto sobre la conciencia? Como me dijo con tono jocoso el famoso psicólogo experimental William Estes después de que me presentaran a él como el hombre que había descubierto el fenómeno del cerebro escindido: «¡Genial, ahora hay dos cosas que no entendemos!». Sin embargo, el interés por ese rompecabezas ya no me abandonaría, como la idea de Polanyi de que la lista de las partes no nos dice cómo funciona el todo. Ambas realidades, la lista de las partes y el modo como dichas partes trabajan juntas para producir su función, exigen una explicación más compleja acerca de cómo esos hechos arrojan luz sobre el problema de la conciencia.

En los últimos treinta años se han invertido miles de millones de dólares en el estudio del papel de las diversas regiones cerebrales y del modo como estas se hallan interconectadas. Sin embargo, ese enfoque local no nos

ofrecerá una explicación exhaustiva de la conciencia, por más que los modernos estudios del cerebro nos digan qué áreas anatómicas específicas están relacionadas con determinadas capacidades mentales. Aunque dichos estudios se añadan a la plétora de datos conocidos sobre el cerebro, ni explican ni explicarán los procesos que este realiza y que, entre otras cosas, se traducen en la conciencia. Si bien el enfoque estructura-función proporciona un conocimiento profundo de cómo el cerebro compartimenta sus múltiples especializaciones, no logra explicar de forma adecuada cómo las reacciones electroquímicas se transforman en experiencias vitales. Ya hemos visto que la estructura y la función son propiedades complementarias: ninguna de ellas nos dice nada sobre la otra. Si no tenemos ni idea de cuál es la función de una neurona, no la deduciremos observando una. Y lo mismo sucede a la inversa. Saber cuál es la función de una neurona no nos dará la menor idea sobre su aspecto. Sin ningún conocimiento previo, la función de las neuronas no puede derivarse de su estructura, ni su estructura puede derivarse de su función. Son dos capas independientes e irreducibles la una a la otra, con protocolos distintos.

La aventura de aprender más sobre las partes subyacentes del cerebro tiene que ampliar su agenda y centrarse también en el diseño neuronal. Limitarse simplemente a tratar de localizar la estructura que produce la conciencia, como intentaron Descartes y muchos de sus predecesores, no nos conducirá al Santo Grial, puesto que la conciencia es inherente *a todo* el cerebro. Cortar grandes trozos de la corteza no interrumpe la conciencia: solo altera su contenido. Esta no está compartimentada en el cerebro como muchas otras capacidades mentales, pongamos por caso la producción del habla o el procesamiento visual, sino que constituye un elemento crucial de todas esas capacidades. Una vez más, como ya hemos visto, la evidencia más convincente en favor de una conciencia fragmentada se revela en la mente de los pacientes con cerebro escindido: cuando se corta la transmisión entre los dos

hemisferios, cada uno de ellos sigue teniendo su propia experiencia consciente.

Aunque la idea de que nuestra conciencia emana de varias fuentes independientes resulta contraria a la intuición, parece ser que tal es el diseño del cerebro. Una vez que este concepto se asimile íntegramente, el verdadero reto será entender cómo los principios de diseño del cerebro permiten que la conciencia surja de ese modo. Este es el futuro desafío para la ciencia que estudia el cerebro.

Unas palabras finales

Cuando empecé este libro, no creía que terminara formulando algunas de las ideas que finalmente he perfilado. La cuestión de fondo siempre ha sido: ¿de verdad la conciencia es un instinto?

En su obra —hoy clásica— *El instinto del lenguaje*, Steven Pinker hace una obligada llamada de atención a la comunidad científica: ¿cómo pueden la mente y el cerebro estar configurados biológicamente y a la vez ser modificados por la experiencia? El libro proporcionaba un marco necesario para reflexionar sobre los límites del aprendizaje y la realidad de las diversas partes de la mente derivadas de la selección natural. Pinker observaba asimismo con acierto que conceptualizar como «instintos» determinados rasgos humanos de orden superior (como el lenguaje) resulta absolutamente discordante.

Incluir el fenómeno de la conciencia en el saco de los instintos —al lado de la ira, la timidez, el afecto, los celos, la envidia, la rivalidad, la sociabilidad, etc.— no es menos desconcertante. Los instintos, como todos sabemos, evolucionan gradualmente, ayudándonos a adaptarnos a nuestro entorno. Añadir la conciencia a la lista de instintos sugiere que esta preciosa propiedad humana, que tanto apreciamos, no es una parte milagrosamente dotada del hardware especial de nuestra especie. Admitir que la conciencia es un instinto

equivale a arrojarla al vasto mundo de la biología, con toda su historia, riqueza, variedad y continuidad. ¿De dónde surgió? ¿Cómo evolucionó? ¿Qué otras especies comparten con nosotros uno u otro rasgo de ella?

Detengámonos un momento para plantearnos la cuestión fundamental: para empezar, ¿qué es un instinto? Este término se utiliza tan profusamente como si fuera confeti en una cabalgata. Año tras año, la lista de instintos no para de crecer. Uno casi pensaría que, si se volara la cabeza, vería un montón de hilos etiquetados, cada uno de los cuales representaría uno de los tan cacareados instintos. De hecho, el cerebro humano «debería» ser una maraña de cables conectados de la forma adecuada para hacer su trabajo; sin embargo, si le pedimos a un neurocientífico que nos muestre cuál es la red correspondiente a un determinado instinto, como la rivalidad o la sociabilidad, nos dirá que no poseemos ese conocimiento; al menos, no todavía. Entonces, ¿en qué ayuda calificar algo como instinto?

Cuando uno se siente confuso con respecto a las definiciones y significados relacionados con el tema de la relación mente/cerebro, siempre resulta reconfortante acudir de nuevo a William James. Hace más de ciento veinticinco años, James escribió un trascendental artículo titulado «¿Qué es un instinto?», donde se apresuraba a definir el concepto:

El instinto generalmente se define como la facultad de actuar de un modo tal que produce ciertos fines, sin previsión de dichos fines, y sin un aprendizaje previo en la ejecución. [...] [Los instintos] son los correlatos funcionales de la estructura. Se puede decir que a la presencia de un determinado órgano casi siempre le acompaña una aptitud innata para su uso. «¿Tiene el ave una glándula para la secreción de aceite? Pues ella sabe instintivamente cómo presionar la glándula para hacer salir el aceite y aplicarlo al plumaje.»⁵

Esta definición parece bastante sencilla y, sin embargo, resulta ingeniosamente dualista. Un instinto necesita una estructura física para funcionar; ahora bien, utilizar dicha estructura requiere de una «aptitud» que, al parecer, se presenta de manera espontánea. Encontrar los correlatos funcionales del aparato físico de un instinto es factible, pero ¿cómo sabemos

cómo se llega a usar? ¿Simplemente ocurre? No parece una respuesta muy científica. ¿El ave empieza con un reflejo innato que le hace presionar la glándula y luego, con el tiempo, aprende que como consecuencia todo funciona mejor? Es evidente que, si no hubiera glándula uropígea, no habría aceite ni oportunidad de aprender a usarlo para volar mejor. Parece pues que el bucle ciego de la selección natural y la experiencia trabajan juntos para configurar lo que llamaríamos un instinto.

Pero estamos hablando del comportamiento de las aves; ¿lo mismo vale para la cognición y la conciencia humanas? James ofrece una explicación de cómo podría funcionar todo esto:

Una sola acción instintiva compleja puede implicar sucesivamente el despertar de [diversos] impulsos. [...] Así, un león hambriento empieza a «buscar» una presa debido a que en él se despierta la imaginación junto con el deseo; empieza a «acecharla» cuando, en el ojo, el oído o la nariz obtiene una impresión de su presencia a cierta distancia; «salta» sobre ella, ya sea cuando la presa se alarma y huye, o cuando la distancia es suficientemente reducida; procede a «desgarrarla» y «devorarla» en el momento en que obtiene una sensación de contacto con sus garras y colmillos. Buscar, acechar, saltar y devorar constituyen otros tantos tipos distintos de contracción muscular, y ninguno de ellos es suscitado por el estímulo apropiado al otro.⁶

Cuando hoy observo el trabajo de James, identifico un esquema en el que encajan los conceptos de módulos y capas. James parece sugerir que los aspectos estructurales de los instintos son módulos innatos incardinados en una arquitectura estratificada. Cada instinto puede funcionar de manera independiente en los comportamientos simples, pero también actúan de manera conjunta. Los instintos individuales pueden secuenciarse de forma coordinada para realizar acciones más complejas que los hacen asemejarse un montón a instintos de orden superior. Esa avalancha de secuencias es lo que llamamos «conciencia». James sostiene que las dinámicas «competitivas» que intervienen en la secuenciación de instintos básicos pueden producir lo que parece ser un comportamiento más complejo manifestado a partir de un estado interno complejo. Incluso añade una descripción de la experiencia del animal

al obedecer un instinto: «Cada impulso y cada paso de cada instinto brilla con su propia luz suficiente, y en ese momento parece la única cosa eternamente correcta y apropiada que hacer. Se hace exclusivamente por sí misma». Eso suena como un conjunto de burbujas unidas por la flecha del tiempo que producen algo así como lo que llamamos «experiencia consciente».

La dinámica que dicta qué burbuja emerge y cuándo lo hace sin duda está influenciada por la experiencia y el aprendizaje. Sin embargo, experiencia, aprendizaje y conciencia deben ser todos ellos elementos isomorfos; esto es, deben operar dentro de un mismo sistema. Cuando se concibe el fenómeno de esta manera, vemos la experiencia consciente como lo que es: un truco de la Madre Naturaleza. Concebir la conciencia como un instinto evolucionado (o una secuencia completa de ellos) nos indica dónde debemos buscar para saber cómo surgió a partir del frío mundo inanimado; nos abre los ojos a la interpretación de que cada aspecto de una experiencia consciente es el despliegue de otros instintos que poseen los humanos y que, por su propia naturaleza, los mecanismos y capacidades que albergan producen el estado sensible de la experiencia consciente. Sorprendentemente, en los últimos años, los biólogos de toda laya han logrado ponerse de acuerdo de manera impresionante para identificar veintinueve redes específicas en el cerebro de la mosca, cada una de las cuales controla un comportamiento concreto. Esos comportamientos individuales pueden combinarse y recombinarse de una forma flexible en patrones más complejos. Pues sí, ¡resulta que las lecciones de la conciencia podemos aprenderlas nada menos que en la mosca de la fruta! Ha empezado la carrera para entender la dimensión física de los instintos.⁷

Sin embargo, muchos aborrecen el uso de conceptos como el de *instinto* para describir la experiencia consciente fenoménica. Para ellos, esa definición también nos despoja a los humanos de nuestro estatus único en el reino animal, a saber, que solo nosotros somos moralmente responsables de nuestros actos. Los humanos podemos decidir hacer y, por lo tanto, podemos decidir «hacer lo correcto». Si la conciencia es un instinto —sostienen—, entonces los humanos

debemos de ser autómatas o zombis estúpidos. Pero si dejamos de lado por el momento las realidades físicas de la mecánica cuántica y los *Schnitts* con su liberador funcionamiento simbólico, podemos argumentar que aceptar la idea de que una entidad compleja como el cerebro/cuerpo/mente tiene un mecanismo cognoscible no nos condena a tales visiones deterministas y desesperadas. El propio James abordó también esta inquietud generalizada:

Aquí cosechamos de inmediato los buenos frutos de nuestra simple concepción fisiológica de lo que es un instinto. Si es un mero impulso excitomotor, debido a la preexistencia de un cierto «arco reflejo» en los centros nerviosos de la criatura, obviamente debe seguir la ley de todos esos arcos reflejos. Una limitación de tales arcos es la de tener su actividad «inhibida» por otros procesos que se desarrollen al mismo tiempo. No importa si el arco se organiza al nacer, o madura espontáneamente más tarde, o se debe al hábito adquirido: debe unir su suerte a la de todos los demás arcos, y a veces tiene éxito, y otras veces falla. [...] La visión mística de un instinto lo haría invariable. El punto de vista fisiológico requiere que muestre irregularidades ocasionales en cualquier animal en el que el número de instintos separados, y la posible entrada del mismo estímulo en varios de ellos, fueran importantes. Y tales irregularidades son las que muestran en abundancia los instintos de todos los animales superiores.⁸

James va mucho más allá, y lo cierto es que se necesita tiempo para asimilar su concepción de los instintos. En cualquier caso, le insto a usted a leer su artículo original para que compruebe por sí mismo la claridad de su pensamiento, la claridad de su escritura y su inquebrantable pragmatismo en estas difíciles cuestiones. James señala el camino a seguir, negándose a aceptar la desesperada caricatura de la humanidad como un robot estrictamente gobernado por respuestas reflejas. Para él, se puede generar un estado conductual complejo variando las combinaciones de módulos simples e independientes, del mismo modo que una combinación de múltiples movimientos pequeños distintos configura el comportamiento complejo de un saltador de pértiga al desplazarse hacia arriba y remontar esta última. Cuando actúan juntos de manera coordinada, hasta los sistemas simples pueden hacer creer al observador que existen otras fuerzas implicadas. James deja clara su postura: «Mi primer acto de libre albedrío será creer en el libre albedrío».

Esta proclamación es coherente con la noción de que las creencias, las ideas y los pensamientos pueden formar parte del sistema mental. Las representaciones simbólicas dentro de este sistema, con toda su flexibilidad y arbitrariedad, están extremadamente ligadas a los mecanismos físicos cerebrales. Las ideas tienen consecuencias, aun en un cerebro físicamente constreñido. No hay por qué desesperar: los estados mentales pueden influir en la acción física en sentido descendente.

En el transcurso de este proyecto, la flexibilidad de mis propias representaciones simbólicas ha sido una fuente de alegría y de sorpresa, no de desesperación. Probablemente el descubrimiento que más me ha sorprendido es el hecho de que hoy haya llegado a creer que los humanos nunca construiremos una máquina que imite nuestra conciencia personal. Las máquinas inanimadas basadas en el silicio funcionan de una manera; y los sistemas vivientes basados en el carbono, de otra muy distinta. Unas lo hacen con un conjunto de instrucciones determinista; los otros, a través de símbolos que conllevan intrínsecamente un cierto grado de incertidumbre.

Esta perspectiva lleva a la conclusión de que la tentativa humana de imitar la inteligencia y la conciencia en las máquinas, un objetivo constante en el ámbito de la inteligencia artificial, está condenada al fracaso. Si los sistemas vivientes funcionan según el principio de complementariedad —la idea de que el lado físico tiene su reflejo en un lado simbólico de carácter arbitrario, cuyos símbolos son resultado de la selección natural—, entonces los modelos puramente deterministas de qué configura la vida siempre se quedarán cortos. En un modelo de inteligencia artificial, la memoria de un acontecimiento está en un lugar único y puede eliminarse con una sola pulsación de tecla. En cambio, en un sistema simbólico viviente y, por ende, estratificado, cada aspecto de un mecanismo puede intercambiarse por otro símbolo siempre que cada uno desempeñe su función adecuada. Es así porque eso es lo que la propia vida permite, de hecho exige: complementariedad.

¿Quién dará forma científica a todas esas ideas? ¿Cómo será la

neurociencia del mañana? En mi opinión, la búsqueda de respuestas duraderas tendrá que incluir a los neuroingenieros, dada su capacidad de alcanzar los principios profundos del diseño de las cosas. Esta revolución todavía se halla en sus comienzos, pero la perspectiva que ofrece es clara. Una arquitectura estratificada, que permite la opción de añadir capas suplementarias, ofrece un marco adecuado para explicar cómo el cerebro se hizo cada vez más complejo a través del proceso de selección natural al tiempo que conservaba las características básicas que habían resultado acertadas. Uno de los retos pendientes es el de identificar lo que hacen las distintas capas de procesamiento, y el mayor de todos es descifrar los protocolos que permiten que una capa interprete los resultados del procesamiento de las capas vecinas. Eso implicará salvar el *Schnitt*, esa brecha epistémica que vincula la experiencia subjetiva con el procesamiento objetivo, y que existe desde que lo hizo la primera célula viviente. El descubrimiento de cómo el lado físico de la brecha, las neuronas, interactúa con el lado simbólico, las dimensiones mentales, solo se logrará mediante el lenguaje de la complementariedad.

En última instancia, debemos entender que la conciencia es un instinto. Forma parte de la vida de los organismos: no necesitamos aprender a producirla ni a utilizarla. En un reciente viaje a Charleston, en Carolina del Sur, mi esposa y yo recorríamos la campiña en busca de un buen pollo frito con pan de maíz. Finalmente encontramos un pequeño restaurante de carretera, nos sentamos y pedimos. Cuando la camarera se alejaba, le dije:

—¡Ah, por cierto! Añada un poco de sémola.

Ella se volvió hacia mí, sonrió y me respondió:

—Cariño, la sémola está incluida.

La sémola está incluida en el pedido, y lo mismo ocurre con eso que llamamos conciencia. Somos afortunados por ambas cosas.

Agradecimientos

Uno de mis chistes favoritos habla de un joven aspirante a estrella de la ópera que canta por primera vez en La Scala de Milán. Tras su debut, el público grita:

—*Ancora!, Ancora!*

Él sonrío para sus adentros y vuelve a cantar un aria a pleno pulmón. De nuevo, el público estalla:

—*Ancora!, Ancora!*

Esto se repite durante cuatro o cinco bises, hasta que finalmente el cantante se dirige al público diciendo:

—Bueno, ya he cantado el aria cinco veces. ¿Qué más queréis?

Entonces, un tipo desde un palco le responde a voz en grito:

—¡Que sigas hasta que la cantes bien!

Cuando completé mi libro anterior —una especie de memorias científicas que contaban la historia de la investigación sobre el cerebro escindido—, creí que sería el último. Aquel fue un libro agradable de escribir porque la perspectiva manaba de la experiencia personal y estaba salpicado de historias que siguen formando una parte importante de mi vida. Pero resultó que aquel libro contenía la semilla que daría origen a otro: el presente volumen. Como me sugirió un lector: «Ahora que ya ha dejado a un lado su historia personal, escriba sobre la conciencia en sí». Esa era una tarea distinta que requería mucho trabajo, un nuevo tipo de trabajo, y mucha ayuda.

Hay una persona que me ha ayudado en este proyecto como ninguna otra: mi hermana Rebecca, que es en parte doctora, en parte botánica, en parte escritora e investigadora científica, y *bon vivant* a tiempo completo. Todos y

todo lo que toca se convierten en algo mejor de lo que eran. Empezó a trabajar conmigo en mis diversos libros justo después de una importante operación quirúrgica a la que tuve que someterme en 2006; en principio, para ayudarme con los aspectos editoriales. No tardó en sentirse fascinada por la neurociencia y muy pronto se convirtió también en mi ayudante de investigación. Su claridad intelectual, su pasión por saber cada vez más sobre todos los temas y su buen humor han tenido un papel fundamental en todo lo que he hecho desde entonces, y sigo estando en deuda con ella.

Cuando Bridget Queenan llegó a mi universidad para impulsar nuestra iniciativa en materia de investigación del cerebro, yo sabía que la rutina cotidiana de la vida académica iba a cambiar para mejor. Su inquebrantable ingenio, empuje e inteligencia, junto con sus sensacionales aptitudes editoriales, supusieron una gran aportación a nuestro esfuerzo. También hubo otras. Obviamente, siempre trato de convertir mis seminarios de posgrado en escenarios de exploración para mis proyectos. Así, el primer año, los estudiantes aportan nuevas ideas materiales, y en esta iniciativa concreta hubo uno en particular, Evan Layer, que resultó de gran ayuda. El segundo año los nuevos alumnos actúan como críticos, editores y varias cosas más para los capítulos en desarrollo.

Con los años, uno forma un pequeño grupo de amistades entre sus colegas de profesión que siempre están dispuestas a leer los diversos borradores de los libros y a comentarlos con detalle, como auténticos amigos, sin escatimar las críticas. En ese sentido, estoy en deuda con Walter Sinnott-Armstrong, que constantemente intentó que no me saliera de los límites de mi pensamiento filosófico, así como con Michael Posner, Steven Hillyard, Leo Chalupa, John Doyle, Marcus Raichle y muchos otros que se esforzaron en ayudarme a mantener encarrilada la historia del cerebro. Y, por último, estoy en deuda también con mi esposa, Charlotte, que es quien de verdad me lleva por el buen camino. Su influencia está en todas partes.

Una vez superados todos los controles internos, el libro se envió a Nueva

York para su publicación. Este ha sido mi primer libro publicado por FSG y espero que no sea el último. Allí, los editores Eric Chinski y Laird Gallagher me alentaron y criticaron a la vez con tanta claridad como fuerza. Después de una primera revisión, sus sugerencias me parecieron tan convincentes que les pedí una segunda. Yo ignoraba que ambos tenían formación filosófica y que leían mis textos con algo más que un mero interés general: los leían con conocimiento de causa, y en innumerables ocasiones me ayudaron a clarificarlos. Luego vino la corrección de estilo de Annie Gottlieb, que sometió cada línea del texto al examen de su mente implacable en aras tanto de su precisión como de su comprensión. Estoy en deuda con todos ellos, y debo señalar que también lo están todos aquellos que decidan leer el producto de mis esfuerzos.

Notas

1. Zan Boag, «Searle: It upsets me when I read the nonsense written by my contemporaries», *NewPhilosopher*, 2 (25 de enero de 2014), <<http://www.newphilosopher.com/articles/john-searle-it-upsets-me-when-i-read-the-nonsense-written-by-my-contemporaries/>>.

2. Henri Frankfort y otros, *The Intellectual Adventure of Ancient Man: An Essay of Speculative Thought in the Ancient Near East*, Chicago, University of Chicago Press, 1977.

3. Robert Rosen, *Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life*, Nueva York, Columbia University Press, 1991, pág. 20.

4. René Descartes, *Discurso del método* (1637), Wikipedia, <https://es.wikisource.org/wiki/Discurso_del_método:_Cuarta_parte>.

5. Gary Hatfield, «René Descartes», en *Stanford Encyclopedia of Philosophy Archive*, 2014, <<http://plato.stanford.edu/archives/fall2015/entries/descartes/>>.

6. René Descartes, *The Philosophical Writings of Descartes, vol. 3, The Correspondence* (ed. y trad. John Cottingham, Robert Stoothoff, Dugald Murdoch y Anthony Kenny), Cambridge, Cambridge University Press, 1984, págs. 19-20.

1. John Locke, *Ensayo sobre el entendimiento humano* (1690), Ayuntamiento de Getafe, <https://www.getafe.es/wp-content/uploads/Locke_John-Ensayo_sobre_el_entendimiento_humano.pdf>.

2. David Hume, «A Letter to a Physician» (1734), en John Hill Burton (comp.), *Life and Correspondence of David Hume*, Edimburgo, William Tait, 1846, pág. 35.

3. Robert G. Brown, «Philosophy Is Bullshit: David Hume», en *Axioms as the Basis for All Understanding*, 2003, consultado el 10 de febrero de 2016 en <<https://www.phy.duke.edu/~rgb/Beowulf/axioms/axioms/node4.html>>.

4. David Hume, *Tratado de la naturaleza humana* (1739-1740), Libros en la Red, Servicio de Publicaciones de la Diputación de Albacete, 2001, pág. 16, <<https://www.dipualba.es/publicaciones/LibrosPapel/LibrosRed/Clasicos/Libros/Hume.pmf>>

5. *Ibid.*, pág. 22.

6. *Ibid.*, pág. 26.

7. David Hume, *Investigación sobre el conocimiento humano* (1748), Madrid, Alianza, 1988, pág. 49.

8. David Hume a John Stewart (1754), carta n.º 1, en *The Letters of David Hume*, vol. 1, 1727-1765, ed. J.Y.T. Greig, 1932; Oxford University Press, 2011, pág. 187.

9. Hume, *Tratado de la naturaleza humana*, *op. cit.*, pág. 191.

10. *Ibid.*, pág. 192.

11. *Ibid.*, pág. 191.

12. Arthur Schopenhauer, *Parerga y paralipómena* (1851); cita: *Essays and Aphorisms*, 1851; reed. Londres, Penguin Group, 2004, pág. 223.

13. Arthur Schopenhauer, *El mundo como voluntad y representación* (1819), vol. II, Madrid, Trotta, 2009, pág. 247.

14. *Ibid.*, págs. 414-415.

15. Cubie King y David Von Drehle, «Encounters with the Arch-Genius, David Gelernter», Time, 25 de febrero de 2016, <<http://time.com/4236974/encounters-with-the-archgenius/>>.

16. Schopenhauer, *El mundo como voluntad y representación*, op. cit., pág. 171.

17. Hermann von Helmholtz, *Tratado de óptica fisiológica* (1867); cita: *Treatise on Physiological Optics*, 1924; reed. Nueva York, Dover, 1962, 2005, vol. 3.

18. Henry Maudsley, *The Physiology and Pathology of Mind*, Nueva York, D. Appleton & Co., 1867, pág. 15, <<https://archive.org/stream/physiologypathol00maudiala#page/14/mode/2up/search/uncons>

19. *Ibid.*, 120.

20. Francis Galton, «Psychometric Experiments», *Brain*, 2 (1879), págs. 149-162.

21. Owen Flanagan, *The Science of the Mind*, Cambridge (Massachusetts), MIT Press, 1984, pág. 60.

22. Franz Brentano, *Psicología desde un punto de vista empírico* (1874); cita: *Psychology from an Empirical Standpoint*, Londres y Nueva York, Routledge, 1995, pág. 68.

23. Flanagan, *Science of the Mind*, *op. cit.*, pág. 62.

24. Drew Westen, «The Scientific Legacy of Sigmund Freud: Toward a Psychodynamically Informed Psychological Science», *Psychological Bulletin*, 124, 1998, pág. 333.

25. Charles Darwin, *El origen de las especies por medio de la selección natural*, tomo III, trad. Antonio de Zulueta, Madrid, Espasa Calpe, 1921, pág. 136.

26. Darwin, *El origen del hombre*, 1871; cita: *The Descent of Man and Selection in Relation to Sex*, en Hutchins, Adler y Brockway, *Great Books of the Western World*, vol. 49, Darwin, pág. 319.

27. Darwin, *El origen de las especies, op. cit.*, pág. 138.

1. William James, *Pragmatismo. Un nuevo nombre para viejas formas de pensar* (1907), Madrid, Alianza, 2000, págs. 57-58.

2. Una pequeña población montañosa de Colorado antaño célebre por sus minas de oro.
(*N. del t.*)

3. *Ibid.*, págs. 61-62.

4. *Ibid.*, pág. 62.

5. *Ibid.*, págs. 79-80.

6. Michael I. Posner, *Chronometric Explorations of Mind*, Hillsdale (Nueva Jersey), Lawrence Erlbaum Associates, 1978.

7. La presencia o ausencia de esas representaciones es la causa de fenómenos descritos en fecha más reciente como el miembro fantasma (presencia de representación, pero ausencia del miembro) o el trastorno de identidad de la integridad corporal (ausencia de representación del miembro, pero presencia de este, que el paciente ve como extraño y quiere que se le ampute).

8. El diencefalo contiene el epitálamo, el tálamo, el hipotálamo, el tálamo ventral y el tercer ventrículo.

9. Wilder Penfield, «Speech, Perception and the Uncommitted Cortex», en John C. Eccles (comp.), *Brain and Conscious Experience*, Nueva York, Springer-Verlag, 1966, pág. 234.

10. *Ibid.*, pág. 235.

11. En inglés, «*Cells that fire together wire together*», que, de manera simplificada, significa que cuando dos neuronas se comunican con frecuencia la interconexión entre ellas se refuerza, lo cual posibilita el aprendizaje mediante la práctica repetitiva. (*N. del t.*)

12. George A. Miller, Psychology: *The Science of Mental Life*, Nueva York, Harper and Row, 1962, pág. 25.

13. David R. Curtis y Per Andersen, «Sir John Carew Eccles, A. C. 27 January 1903—2 May 1997», *Biographical Memoirs of Fellows of the Royal Society*, 47, 2001, págs. 160-187, <<https://www.science.org.au/fellowship/fellows/biographical-memoirs/john-carew-eccles-1903-1997#2>>.

14. Karl R. Popper y John C. Eccles, *The Self and Its Brain: An Argument for Interactionism*, Berlín, Springer-Verlag, 1977.

15. Henry H. Dale, «The Beginnings and the Prospects of Neurohumoral Transmission», *Pharmacological Reviews*, 6, 1954, págs. 7-13.

16. John C. Eccles, «Hypotheses Relating to the Brain-Mind Problem», *Nature*, 168, 1951, págs. 53-57.

17. E. G. Walsh, «Brain and Conscious Experience: Study Week September 28 to October 4, 1964 of the Pontificia Academia Scientiarum. Edited by Sir John C. Eccles. Berlin, Heidelberg, New York: Springer-Verlag...» (reseña bibliográfica), *Quarterly Journal of Experimental Physiology and Cognate Medical Sciences*, 52, 1967, pág. 330.

18. William H. Thorpe, «Ethology and Consciousness», en *Eccles, Brain and Conscious Experience, op. cit.*, pág. 44.

19. John C. Eccles, «Conscious Experience and Memory», en *Eccles, Brain and Conscious Experience*, *op. cit.*, pág. 326.

20. Una operación quirúrgica que consiste en seccionar el cuerpo calloso del cerebro y que se utiliza como último recurso para tratar ciertos casos de epilepsia resistentes a los fármacos. (*N. del t.*)

21. Roger W. Sperry, «Brain Bisection and Mechanisms of Consciousness», en *Eccles, Brain and Conscious Experience, op. cit.*, pág. 299.

22. Roger W. Sperry, «Mind-Brain Interaction: Mentalism, Yes; Dualism, No», *Neuroscience*, 5, 1980, pág. 196.

23. Sperry, «Brain Bisection», *op. cit.*, pág. 308.

24. *Ibid.*

25. Eccles, *Brain and Conscious Experience*, *op. cit.*, pág. 250.

26. *Ibid.*, pág. 248.

27. Charles G. Gross, «Hans-Lukas Teuber: A Tribute», *Cerebral Cortex*, 4, 1994, págs. 451-454.

28. Eccles, *Brain and Conscious Experience*, *op. cit.*, pág. 582.

29. Roger W. Sperry, «Mind, Brain, and Humanist Values», *Bulletin of the Atomic Scientists*, 22, 1966, págs. 2-6.

30. Roger W. Sperry, «Perception in the Absence of the Neocortical Commissures», en David A. Hamburg, Karl H. Pribram y Albert J. Stunkard (comps.), *Perception and Its Disorders*, vol. 48, Baltimore, Williams and Wilkins, 1970, págs. 123-128.

31. Donald M. MacKay, «Soul, Brain Science and the entry», en R. L. Gregory (comp.), *The Oxford Companion to the Mind*, Oxford, Oxford University Press, 1987, págs. 724-725.

32. La cursiva es mía; creo que ayuda a clarificar la frase. (*N. del t.*)

33. En inglés *raw feel*, para diferenciarla de la *cooked feel* o «sensación elaborada». (N. del t.)

34. M. S. Hacker, «The Sad and Sorry History of Consciousness: Being, among Other Things, a Challenge to the “Consciousness-Studies Community”», *Royal Institute of Philosophy Supplement*, 70, 2012, págs. 149-168.

35. Thomas Nagel, «The Psychophysical Nexus», en Paul Boghossian y Christopher Peacocke (comps.), *New Essays on the A Priori*, Oxford, Oxford University Press, 2000, págs. 432-471.

36. Douglas R. Hofstadter y Daniel C. Dennett, *The Mind's I: Fantasies and Reflections on Self and Soul*, Nueva York, Basic Books, 1981, 2000, pág. 409.

37. Owen Flanagan, *The Problem of the Soul: Two Visions of Mind and How to Reconcile Them*, Nueva York, Basic Books, 2002, pág. 87.

38. Francis H. Crick, «Thinking About the Brain», *Scientific American*, 241, 1979, págs. 219-232.

39. Michael I. Posner y Mary K. Rothbart, «Attentional Mechanisms and Conscious Experience», en A.D. Milner y M.D. Rugg (comps.), *The Neuropsychology of Conscious Experience*, Londres, Academic Press, 1992, págs. 97-117.

40. Michael S. Gazzaniga, *The Bisected Brain*, Nueva York, Appleton Century Crofts, 1970.

41. Crick, «Thinking About the Brain», *op. cit.*

42. *Ibid.*

43. *Ibid.*

44. Francis Crick y Christof Koch, «Towards a Neurobiological Theory of Consciousness», *Seminars in the Neurosciences*, 2, 1990, págs. 263-275.

45. Christof Koch, *The Quest for Consciousness: A Neurobiological Approach*, Englewood (CO), Roberts and Company, 2004, pág. 17.

46. *Ibid.*, pág. 15.

1. Charles S. Sherrington, *Man on His Nature: The Gifford Lectures, 1937-38, 1940*; reed. Cambridge, Cambridge University Press, 2009.

2. Con esta metáfora, Sherrington comparaba el cerebro con el telar mecánico de Jacquard, que en la época en la que escribía era el dispositivo mecánico más complejo inventado hasta la fecha. (*N. del t.*)

3. Michael S. Gazzaniga, «Brain Mechanisms and Conscious Experience», *Experimental and Theoretical Studies of Consciousness*, CIBA Foundation Symposium, 174, Chichester, John Wiley and Sons, 1993, págs. 247-262.

4. Edoardo Bisiach y Claudio Luzzatti, «Unilateral Neglect of Representational Space», *Cortex*, 14, 1978, págs. 129-133.

5. Patrik Vuilleumier, «Mapping the Functional Neuroanatomy of Spatial Neglect and Human Parietal Lobe Functions: Progress and Challenges», *Annals of the New York Academy of Sciences*, 1296, 2013, págs. 50-74.

6. Bruce T. Volpe, Joseph E. Ledoux y Michael Gazzaniga, «Information Processing of Visual Stimuli in an “Extinguished” Field», *Nature*, 282, 1979, págs. 722-724.

7. El explorador angloirlandés Ernest Shackleton fue el primero en describir esta sensación de una presencia extraña. Él y sus dos compañeros se hallaban en un estado de absoluto agotamiento y privación física después de atravesar con éxito 680 millas de los mares más procelosos del mundo en un bote salvavidas con agujeros y sin apenas comida y agua. Se encontraban en la última etapa de una misión épica en busca de ayuda para su tripulación, varada en una isla en la costa de la Antártida: cruzar dos cadenas montañosas inexploradas y cubiertas de nieve en la isla Georgia del Sur, con solo un piolet y unos 14 metros de cuerda, lo más rápido posible. Durante esta caminata, Shackleton describió la sensación de que les acompañaba un cuarto hombre. Más tarde, T. S. Eliot haría alusión a ese mismo fenómeno en su poema *La tierra baldía*; pero él lo denominó «el tercer hombre», y el nombre cuajó (véase John Geiger, *El tercer hombre: el secreto de la supervivencia en entornos extremos*, Barcelona, Ariel, 2009).

8. Reinhold Messner, *The Naked Mountain*, Seattle, The Mountaineers Books, 2003, pág. 299.

9. William Dewi Rees, «The Hallucinations of Widowhood», *British Medical Journal*, 4, 1971, pág. 37.

10. Shahar Arzy y otros, «Induction of an Illusory Shadow Person», *Nature*, 443, 2006, pág. 287.

11. Olaf Blanke y otros, «Neurological and Robot-Controlled Induction of an Apparition», *Current Biology*, 24, 2014, págs. 2.681-2.686.

12. *Ibid.*

13. Frederico A. C. Azevedo y otros, «Equal Numbers of Neuronal and Nonneuronal Cells Make the Human Brain an Isometrically Scaled-up Primate Brain», *Journal of Comparative Neurology*, 513, 2009, págs. 532-541.

14. Suzana Herculano-Houzel, «The Human Brain in Numbers: A Linearly Scaled-up Primate Brain», *Frontiers in Human Neuroscience*, 3, 2009, pág. 31.

15. Mark E. Nelson y James M. Bower, «Brain Maps and Parallel Computers», *Trends in Neurosciences*, 13, 1990, págs. 403-408.

16. Donald D. Clarke y Louis Sokoloff, «Circulation and Energy Metabolism of the Brain», en George J. Siegel y otros (comps.), *Basic Neurochemistry: Molecular, Cellular and Medical Aspects*, Filadelfia, Lippincott-Raven, 1999, págs. 637-670.

17. Georg F. Striedter, *Principles of Brain Evolution*, Sunderland (Massachusetts), Sinauer Associates, 2005.

18. David Meunier, Renaud Lambiotte y Edward T. Bullmore, «Modular and Hierarchically Modular Organization of Brain Networks», *Frontiers in Neuroscience*, 4, 2010, pág. 200.

19. *Ibid.*

20. Dmitri B. Chklovskii, Thomas Schikorski y Charles F. Stevens, «Wiring Optimization in Cortical Circuits», *Neuron*, 34, 2002, págs. 341-347.

21. Danielle S. Bassett y otros, «Dynamic Reconfiguration of Human Brain Networks during Learning», *Proceedings of the National Academy of Sciences (PNAS)*, 108, 2011, págs. 7.641-7.646; Danielle S. Bassett y otros, «Robust Detection of Dynamic Community Structure in Networks», *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 23, 2013, 013142.

22. Olaf Sporns y Richard F. Betzel, «Modular Brain Networks», *Annual Review of Psychology*, 67, 2016, págs. 613-640.

23. Striedter, *Principles of Brain Evolution*, *op. cit.*, pág. 248.

24. Beth L. Chen, David H. Hall y Dmitri B. Chklovskii, «Wiring Optimization Can Relate Neuronal Structure and Function», PNAS, 103, 2006, 4.723-4.728; Christopher Cherniak y otros, «Global Optimization of Cerebral Cortex Layout», PNAS, 101, 2004, págs. 1.081-1.086; Yong-Yeol Ahn, Hawoong Jeong y Beom Jun Kim, «Wiring Cost in the Organization of a Biological Neuronal Network», *Physica A: Statistical Mechanics and Its Applications*, 367, 2006, págs. 531-537.

25. Jef Clune, Jean-Baptiste Mouret y Hod Lipson, «The Evolutionary Origins of Modularity», *Proceedings of the Royal Society of London B: Biological Sciences*, 280, 2013, 2012.2863.

26. Peter Carruthers, *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought*, Oxford, Oxford University Press, 2006.

27. Sporns y Betzel, «Modular Brain Networks», *op. cit.*

28. Nicola Clayton y Nathan Emery, «Corvid Cognition», *Current Biology*, 15, 2005, págs. R80-R81.

29. Alex H. Taylor y otros, «Complex Cognition and Behavioural Innovation in New Caledonian Crows», *Proceedings of the Royal Society of London. B: Biological Sciences*, 277, 2010, págs. 2.637-2.643.

30. Jennifer C. Holzhaider, Gavin R. Hunt y Russell D. Gray, «Social Learning in New Caledonian Crows», *Learning and Behavior*, 38, 2010, págs. 206-219.

31. Gavin R. Hunt, C. Lambert y Russell D. Gray, «Cognitive Requirements for Tool Use by New Caledonian Crows (*Corvus moneduloides*)», *New Zealand Journal of Zoology*, 34, 2007, págs. 1-7.

32. Andrew Whiten y otros, «Emulation, Imitation, Over-Imitation and the Scope of Culture for Child and Chimpanzee», *Philosophical Transactions of the Royal Society of London. B: Biological Sciences*, 364, 2009, págs. 2.417-2.428.

33. Wolfgang Köhler, *The Mentality of Apes*, Londres, Kegan Paul, Trench, Trübner & Co., 1925.

34. Kristin Liebal y otros, «Infants Use Shared Experience to Interpret Pointing Gestures», *Developmental Science*, 12, 2009, págs. 264-271.

35. David Premack, «Why Humans Are Unique: Three Theories», *Perspectives on Psychological Science*, 5, 2010, págs. 22-32.

36. Carruthers, *Architecture of the Mind*, *op. cit.*

37. David Premack y Guy Woodruff, «Does the Chimpanzee Have a Theory of Mind?», *Behavioral and Brain Sciences*, 1, 1978, págs. 515-526.

38. Josep Call y Michael Tomasello, «Does the Chimpanzee Have a Theory of Mind?» 30 Years Later», *Trends in Cognitive Sciences*, 12, 2008, págs. 187-192.

39. Zijing He, Matthias Bolz y Renée Baillargeon, «Understanding of False Belief in 2.5-year-olds in a Violation-of-Expectation Test», ponencia presentada en la Reunión Bianual de la Sociedad para la Investigación del Desarrollo Infantil, Boston, marzo de 2007.

40. Christopher Krupenye y otros, «Great Apes Anticipate That Other Individuals Will Act According to False Beliefs», *Science*, 354, 2016, págs. 110-114.

41. John W. Pilley y Alliston K. Reid, «Border Collie Comprehends Object Names as Verbal Referents», *Behavioural Processes*, 86, 2011, págs. 184-195; John W. Pilley, «Border Collie Comprehends Sentences Containing a Prepositional Object, Verb, and Direct Object», *Learning and Motivation*, 44, 2013, págs. 229-240.

42. Katharina C. Kirchofer y otros, «Dogs (*Canis familiaris*), but Not Chimpanzees (*Pan troglodytes*), Understand Imperative Pointing», *PloS One*, 7, 2012, e30913.

43. Michelle E. Magininity y Randolph C. Grace, «Visual Perspective Taking by Dogs (Canis familiaris) in a Guesser-Knower Task: Evidence for a Canine Theory of Mind?», *Animal Cognition*, 17, 2014, págs. 1.375-1.392.

44. Brian Hare y Michael Tomasello, «Human-like Social Skills in Dogs?», *Trends in Cognitive Sciences*, 9, 2005, págs. 439-444.

45. Muhammad A. Spocter y otros, «Neuropil Distribution in the Cerebral Cortex Differs between Humans and Chimpanzees», *Journal of Comparative Neurology*, 520, 2012, págs. 2.917-2.929.

46. Julia Mehlhorn y otros, «Tool-Making New Caledonian Crows Have Large Associative Brain Areas», *Brain, Behavior and Evolution*, 75, 2010, págs. 63-70.

47. Justin S. Feinstein y otros, «The Human Amygdala and the Induction and Experience of Fear», *Current Biology*, 21, 2011, págs. 34-38.

1. Robert Rosen, *Dynamical System Theory in Biology*, Nueva York, Wiley, 1970.

2. Michael Polanyi, «Life's Irreducible Structure», *Science*, 160, 1968, pág. 1.308.

3. *Ibid.*

4. Marie E. Csete y John C. Doyle, «Reverse Engineering of Biological Complexity», *Science*, 295, 2002, págs. 1.664-1.669.

5. John C. Doyle y Marie E. Csete, «Architecture, Constraints, and Behavior», PNAS, 108, Suplemento 3, 2011, págs. 15.624-15.630.

6. Polanyi, «Life's Irreducible Structure», *op. cit.*

7. Doyle y Csete, «Architecture, Constraints, and Behavior», *op. cit.*

8. David L. Alderson y John C. Doyle, «Contrasting Views of Complexity and Their Implications for Network-Centric Infrastructures», *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans*, 40, 2010, pág. 840.

9. *Ibid.*

10. Doyle y Csete, «Architecture, Constraints, and Behavior», *op. cit.*

11. Jerzy Wegiel y otros, «The Neuropathology of Autism: Defects of Neurogenesis and Neuronal Migration, and Dysplastic Changes», *Acta Neuropathologica*, 119, 2010, págs. 755-770.

12. Aswin Sekar y otros, «Schizophrenia Risk from Complex Variation of Complement Component 4», *Nature*, 530, 2016, págs. 177-183.

13. Alderson y Doyle, «Contrasting Views of Complexity», *op. cit.*

14. Un amigo mío que leyó esto me preguntó: «¿Qué es una pérgola? No es precisamente un término que los tíos del Lower East Side neoyorquino hayamos oído mucho». Si al lector le ocurre lo mismo, voy a dejar que lo busque en Wikipedia, tal como terminó haciendo mi amigo. De esa forma podrá ver también unas cuantas imágenes.

15. Mung Chiang y otros, «Layering as Optimization Decomposition: A Mathematical Theory of Network Architectures», *Proceedings of the IEEE*, 95, 2007, págs. 255-312.

16. En español, la palabra *subsunción* suele definirse simplemente como «Acción y afecto de subsumir»; a su vez, *subsumir* se define como «Incluir algo como componente en una síntesis o clasificación más abarcadora», «Considerar algo como parte de un conjunto más amplio o como caso particular sometido a un principio o norma general» (RAE), o simplemente «Incluir algo en un conjunto más amplio» (*Diccionario de uso del español*). (N. del t.)

17. Harold Pashler, *Encyclopedia of the Mind, vol. 1*, Thousand Oaks (California), Sage Publications, 2013, pág. 465.

18. Tony J. Prescott, Peter Redgrave y Kevin Gurney, «Layered Control Architectures in Robots and Vertebrates», *Adaptive Behavior*, 7, 1999, págs. 99-127.

19. *Ibid.*, pág. 101.

20. Marc Kirschner y John Gerhart, «Evolvability», PNAS, 95, 1998, págs. 8.420-8.427.

21. Peter T. Boag y Peter R. Grant, «Intense Natural Selection in a Population of Darwin's Finches (Geospizinae) in the Galápagos», *Science*, 214, 1981, págs. 82-85.

22. John Gerhart y Marc Kirschner, «The Theory of Facilitated Variation», PNAS, 104, Suplemento 1, 2007, págs. 8.582-8.589.

23. Alderson y Doyle, «Contrasting Views of Complexity», *op. cit.*

24. Doyle y Csete, «Architecture, Constraints, and Behavior», *op. cit.*

25. *Ibid.*

26. Christopher W. Johnson, «What Are Emergent Properties and How Do They Afect the Engineering of Complex Systems?», *Reliability Engineering and System Safety*, 91, 2006, págs. 1.475-1.481.

27. Eve Marder, «Variability, Compensation and Modulation in Neurons and Circuits», PNAS, 108, Suplemento 3, 2011, págs. 15.542-15.548.

28. Tamar Friedlander y otros, «Evolution of Bow-Tie Architectures in Biology», *PLoS Computational Biology*, 11, 2015, e1004055.

29. John C. Doyle, «Guaranteed Margins for LQG Regulators», IEEE Transactions on Automatic Control, 23, 1978, págs. 756-757.

30. Alderson y Doyle, «Contrasting Views of Complexity», *op. cit.*

31. Arne J. Nagengast, Daniel A. Braun y Daniel M. Wolpert, «Risk-Sensitive Optimal Feedback Control Accounts for Sensorimotor Behavior Under Uncertainty», *PLoS Computational Biology*, 6, 2010, e1000857.

32. Fiona A. Chandra, Gentian Buzi y John C. Doyle, «Glycolytic Oscillations and Limits on Robust Efficiency», *Science*, 333, 2011, págs. 187-192.

33. Daniel Kahneman, Pensar rápido, pensar despacio, Barcelona, *Debate*, 2015.

34. Roger W. Sperry, «Neurology and the Mind-Brain Problem», *American Scientist*, 40, 1952, págs. 291-312.

35. Véase el vídeo «Daniel Wolpert: La verdadera razón del cerebro», en inglés con subtítulos y transcripción en español, en https://www.ted.com/talks/daniel_wolpert_the_real_reason_for_brains/transcript?language=es.

1. Los procesos necesarios para la conciencia se inician en la parte evolutivamente más antigua del cerebro: el tronco del encéfalo o tallo cerebral. La principal función del tallo cerebral es la regulación homeostática del cuerpo y el cerebro. Es lo que hace que el corazón lata, los pulmones respiren y los intestinos digieran. Si se desconecta el tallo cerebral de cualquier mamífero, su cuerpo muere. Desde el tallo cerebral parten neuronas en muchas direcciones; las necesarias para la conciencia están conectadas con los núcleos intralaminares del tálamo, situado entre el mesencéfalo y la corteza.

2. David A. Drachman, «The Amyloid Hypothesis, Time to Move On: Amyloid Is the Downstream Result, Not Cause, of Alzheimer's Disease», *Alzheimer's and Dementia*, 10, 2014, págs. 372-380; Jessica Freiherr y otros, «Intranasal Insulin as a Treatment for Alzheimer's Disease: A Review of Basic Research and Clinical Evidence», *CNS Drugs*, 27, 2013, págs. 505-514.

3. Stanley B. Klein, Leda Cosmides y Kristi A. Costabile, «Preserved Knowledge of Self in a Case of Alzheimer's Dementia», *Social Cognition*, 21, 2003, págs. 157-165.

4. Lydia Krabbendam y Jim van Os, «Schizophrenia and Urbanicity: A Major Environmental Influence—Conditional on Genetic Risk», *Schizophrenia Bulletin*, 31, 2005, págs. 795-799.

5. Elizabeth Cantor-Graae y Jean-Paul Selten, «Schizophrenia and Migration: A Meta-Analysis and Review», *American Journal of Psychiatry*, 162, 2005, págs. 12-24.

6. Wim Velting y otros, «Ethnic Density of Neighborhoods and Incidence of Psychotic Disorders among Immigrants», *American Journal of Psychiatry*, 165, 2008, págs. 66-73.

7. Theresa H. M. Moore y otros, «Cannabis Use and Risk of Psychotic or Affective Mental Health Outcomes: A Systematic Review», *The Lancet*, 370, 2007, págs. 319-328; Robin M. Murray y otros, «Cannabis, the Mind and Society: The Hash Realities», *Nature Reviews Neuroscience*, 8, 2007, págs. 885-895.

8. Kurt Schneider, *Psicopatología clínica*, Madrid, Triacastela, 1997.

9. Jim van Os y Shitij Kapur, «Schizophrenia», *The Lancet*, 374 (2009), 635-645; Jim van Os, «“Salience Syndrome” Replaces “Schizophrenia” in DSM-V and ICD-11: Psychiatry’s Evidence-Based Entry into the 21st Century?», *Acta Psychiatrica Scandinavica*, 120, 2009, págs. 363-372.

10. Shitij Kapur, «Psychosis as a State of Aberrant Salience: A Framework Linking Biology, Phenomenology, and Pharmacology in Schizophrenia», *American Journal of Psychiatry*, 160, 2003, págs. 13-23.

11. Marc Laruelle, «Imaging Dopamine Transmission in Schizophrenia: A Review and Meta-Analysis», *Quarterly Journal of Nuclear Medicine*, 42, 1998, págs. 211-221; Oliver Guillin, Anissa Abi-Dar gham y Marc Laruelle, «Neurobiology of Dopamine in Schizophrenia», *International Review of Neurobiology*, 78, 2007, págs. 1-39.

12. Kapur, «Psychosis as a State of Aberrant Saliience», *op. cit.*

13. Jimmy Jensen y otros, «The Formation of Abnormal Associations in Schizophrenia: Neural and Behavioral Evidence», *Neuropsychopharmacology*, 33, 2008, págs. 473-479; J. P. Roiser y otros, «Do Patients with Schizophrenia Exhibit Aberrant Salience?», *Psychological Medicine*, 39, 2009, págs. 199-209.

14. Rosalind Cartwright, «Sleepwalking Violence: A Sleep Disorder, a Legal Dilemma, and a Psychological Challenge», *American Journal of Psychiatry*, 161, 2004, págs. 1.149-1.158.

15. Pierre Maquet y otros, «Functional Neuroanatomy of Human Slow Wave Sleep», *Journal of Neuroscience*, 17, 1997, págs. 2.807-2.812; A. R. Braun y otros, «Regional Cerebral Blood Flow Throughout the Sleep-Wake Cycle: An H215O PET Study», *Brain*, 120, 1997, págs. 1.173-1.197; Jesper L. R. Andersson y otros, «Brain Networks Affected by Synchronized Sleep Visualized by Positron Emission Tomography», *Journal of Cerebral Blood Flow and Metabolism*, 18, 1998, págs. 701-715; C. Kaufmann y otros, «Brain Activation and Hypothalamic Functional Connectivity During Human Non-Rapid Eye Movement Sleep: An EEG/fMRI Study», *Brain*, 129, 2006, págs. 655-667.

16. Claudio Bassetti y otros, «SPECT During Sleepwalking», *The Lancet*, 356, 2000, págs. 484-485.

17. Michele Terzaghi y otros, «Evidence of Dissociated Arousal States during NREM Parasomnia from an Intracerebral Neurophysiological Study», *Sleep*, 32, 2009, págs. 409-412.

18. Steven Laureys y otros, «The Locked-in Syndrome: What Is It Like to Be Conscious but Paralyzed and Voiceless?», *Progress in Brain Research*, 150, 2005, págs. 495-511.

19. Jean-Dominique Bauby, *La escafandra y la mariposa*, Barcelona, Del Bronce, 2007.

20. *Ibid.*

21. Sofiane Ghorbel, «Statut fonctionnel et qualité de vie chez le locked-in syndrome a domicile», DEAMotricité Humaine et Handicap, Saint-Étienne (Montpellier), *Laboratorio de Bioestadística, Epidemiología e Investigación Clínica*, Universidad Jean Monnet, 2002.

22. Ronald E. Cranford, «The Persistent Vegetative State: The Medical Reality (Getting the Facts Straight)», *Hastings Center Reports*, 18, 1988, págs. 27-32.

23. Steven Laureys, Olivia Gosseries y Giulio Tononi (comps.), *The Neurology of Consciousness: Cognitive Neuroscience and Neuropathology*, 2.^a ed., Amsterdam, Academic Press, 2016.

24. Bjorn Merker, «Consciousness Without a Cerebral Cortex: A Challenge for Neuroscience and Medicine», *Behavioral and Brain Sciences*, 30, 2007, págs. 63-81.

25. Jaak Panksepp y otros, «Effects of Neonatal Decortication on the Social Play of Juvenile Rats», *Physiology and Behavior*, 56, 1994, págs. 429-443.

26. Jaak Panksepp, «Affective Consciousness: Core Emotional Feelings in Animals and Humans», *Consciousness and Cognition*, 14, 2005, págs. 30-80.

27. David J. Anderson y Ralph Adolphs, «A Framework for Studying Emotions across Species», *Cell*, 157, 2014, págs. 187-200.

28. Jaak Panksepp, Thomas Fuchs y Paolo Iacobucci, «The Basic Neuroscience of Emotional Experiences in Mammals: The Case of Subcortical FEAR Circuitry and Implications for Clinical Anxiety», *Applied Animal Behaviour Science*, 129, 2011, págs. 1-17; Jaak Panksepp, «Affective Neuroscience of the Emotional BrainMind: Evolutionary Perspectives and Implications for Understanding Depression», *Dialogues in Clinical Neuroscience*, 12, 2010, págs. 533-545.

29. Jaak Panksepp y Jules B. Panksepp, «The Seven Sins of Evolutionary Psychology», *Evolution and Cognition*, 6, 2000, págs. 108-131.

30. Joseph LeDoux, «Rethinking the Emotional Brain», *Neuron*, 73, 2012, págs. 653-676.

31. Steven Pinker, *El instinto del lenguaje*, Barcelona, Alianza Editorial, 2012.

32. Panksepp, «Affective Consciousness», *op. cit.*

33. Bernard J. Baars, *A Cognitive Theory of Consciousness*, Cambridge (Reino Unido), Cambridge University Press, 1988.

34. L. F. Haas, «Phineas Gage and the Science of Brain Localisation», *Journal of Neurology, Neurosurgery, and Psychiatry*, 71, 2001, pág. 761.

35. Sergio Paradiso y otros, «Frontal Lobe Syndrome Reassessed: Comparison of Patients with Lateral or Medial Frontal Brain Damage», *Journal of Neurology, Neurosurgery and Psychiatry*, 67, 1999, págs. 664-667.

1. John Tyndall, *Fragments of Science for Unscientific People: A Series of Detached Essays, Lectures, and Reviews*, vol. 1, Nueva York, D. Appleton & Co., 1871, pág. 119.

2. En inglés *explanatory gap*, también traducido como «vacío explicativo». (*N. del t.*)

3. Joseph Levine, *Purple Haze: The Puzzle of Consciousness*, Oxford, Oxford University Press, 2001, pág. 6.

4. *Ibid.*, pág. 87.

5. David J. Chalmers, «Facing Up to the Problem of Consciousness», *Journal of Consciousness Studies*, 2, 1995, págs. 200-219.

6. John Tyndall, «The Belfast Address», en Tyndall, *Fragments of Science*, vol. 2, Londres, Longmans, Green & Co., 1879, <<http://www.gutenberg.org/files/24527/24527-h/24527-h.htm#Toc158391647>>.

7. William James, *Principios de psicología* (1890); cita: *The Principles of Psychology*, en Hutchins, Adler and Brockway, *Great Books of the Western World*, vol. 53, William James, pág. 97.

8. *Ibid.*, pág. 95.

9. *Ibid.*

10. William Stukeley, *Memoirs of Sir Isaac Newton's Life*, manuscrito, 1752; ed. facsímil, Royal Society, 2010; consultado el 26 de junio de 2016 en <<http://ttp.royalsociety.org/ttp/ttp.html?id=1807da00-909a-4abf-b9c1-0279a08e4bf2&type=book>>.

11. John Conduitt, «Draft account of Newton's life at Cambridge (1727-28)», Keynes Ms. 130.04, Cambridge (Reino Unido), *King's College*, consultado el 26 de junio de 2016 en <<http://www.newtonproject.sussex.ac.uk/catalogue/record/THEM00167>>.

12. Rudolf Clausius, *The Mechanical Theory of Heat - with its Applications to the Steam-engine and to the Physical Properties of Bodies*, Londres, John van Voorst, 1867.

13. Max Planck, «On the Law of Distribution of Energy in the Normal Spectrum», *Annalen der Physik*, 4, 1901, pág. 553.

14. Helge Kragh, «Max Planck: The Reluctant Revolutionary», *Physics World*, 13, 2000, pág. 31.

15. L. Piazza y otros, «Simultaneous Observation of the Quantization and the Interference Pattern of a Plasmonic Near-Field», *Nature Communications*, 6, 2015, pág. 6.407.

16. Richard Feynman, «Sir Douglas Robb Memorial Lecture», Universidad de Auckland, 1979, consultado el 2 de septiembre de 2016 en <<https://www.youtube.com/watch?v=xdZMXWmlp9g>>.

17. Richard Feynman, «“The Quantum View of Physical Nature” (Messenger Lecture)», Universidad Cornell, 1964, consultado el 3 de octubre de 2016 en <<https://www.youtube.com/watch?v=x5RQ3QF9GGI>>.

18. John von Neumann, *Fundamentos matemáticos de la mecánica cuántica* (1932); cita: *Mathematical Foundations of Quantum Mechanics*, Princeton (Nueva Jersey), Princeton University Press, 1955.

19. Feynman, «Messenger Lecture», *op. cit.*

20. Jim Baggott, *La historia del cuanto* (2017); cita: *The Quantum Story: A History in 40 Moments*, Oxford, Oxford University Press, 2011, pág. 100.

21. Obviamente, hay que entender aquí *momento* en su acepción física, esto es, como cantidad de movimiento. (*N. del t.*)

22. *Ibid.*

23. Robert Rosen, «On the Limitations of Scientific Knowledge», en John L. Casti y Anders Karlqvist (comps.), *Boundaries and Barriers: On the Limits to Scientific Knowledge, Reading* (Massachusetts), Perseus Books, 1996, pág. 203.

24. *Ibid.*

1. Howard Hunt Pattee, «Physical and Functional Conditions for Symbols, Codes, and Languages», *Biosemitotics*, 1, 2008, págs. 147-168.

2. Howard Hunt Pattee y Joanna R ą czaszek-Leonardi, *Laws, Language and Life: Howard Pattee's Classic Papers on the Physics of Symbols with Contemporary Commentary*, Dordrecht, Springer, 2012, pág. 7.

3. *Ibid.*, pág. 8.

4. Pattee, «Physical and Functional Conditions for Symbols, Codes, and Languages», *op. cit.*

5. Howard Hunt Pattee, «The Physical Basis of Coding and Reliability in Biological Evolution», en Pattee y Rączaszek-Leonardi, *Laws, Language and Life, op. cit.*, págs. 33-54; reprod. de *Towards a Theoretical Biology 1, Prolegomena, Actas del Simposio de la Unión Internacional de Ciencias Biológicas*, Bellagio, agosto-septiembre de 1966, ed. C. H. Waddington, Edimburgo, Edinburgh University Press, 1968, págs. 67-93.

6. Pattee y Rączaszek-Leonardi, *Laws, Language and Life*, *op. cit.*, pág. 10.

7. Howard Hunt Pattee, «Physical Problems of Decision-Making Constraints», en Pattee y Rączaszek-Leonardi, *Laws, Language and Life*, *op. cit.*, pág. 70; reprod. de *International Journal of Neuroscience*, 3, 1972, págs. 99-106.

8. Pattee, «Physical Basis of Coding and Reliability», *op. cit.*

9. A Pascal le interesaba especialmente la teoría de la probabilidad. Básicamente la apuesta tiene que ver con qué nivel de riesgo es aceptable si la consecuencia de estar equivocado es nefasta. ¿Vale la pena arriesgarse a las sustanciales consecuencias de creer que Dios no existe si resulta que sí existe?

10. John von Neumann, *Theory of Self-Reproducing Automata*, Urbana (Illinois), University of Illinois Press, 1966; Erwin Schrodinger, *What Is Life? The Physical Aspect of the Living Cell*, 1944; reed. Cambridge (Reino Unido), Cambridge University Press, 2012.

11. Steve Martin, <https://www.youtube.com/watch?v=q_8amMzGAx4>.

12. Howard Hunt Pattee, «The Complementarity Principle in Biological and Social Structures», en Pattee y Rączaszek-Leonardi, *Laws, Language and Life, op. cit.*, págs. 143-154; reprod. de *Journal of Social Biology Structures*, 1, 1978, págs. 191-200.

13. Howard Hunt Pattee, «Cell Psychology: An Evolutionary Approach to the Symbol-Matter Problem», en Pattee y Rączaszek-Leonardi, *Laws, Language and Life*, *op. cit.*, pág. 170; reprod. de *Cognition and Brain Theory*, 5, 1982, págs. 325-341.

14. Marcello Barbieri, «Biosemiotics: A New Understanding of Life», *Naturwissenschaften*, 95, 2008, pág. 579.

15. *Ibid.*, pág. 597.

16. *Ibid.*

17. *Ibid.*, pág. 580.

18. *Ibid.*, pág. 596.

19. Noa Liscovitch-Brauer y otros, «Trade-off between Transcriptome Plasticity and Genome Evolution in Cephalopods», *Cell*, 169, 2017, págs. 191-202.

20. Christian B. Anfinsen, «Principles That Govern the Folding of Protein Chains», *Science*, 181, 1973, págs. 223-230.

21. Von Neumann, *Theory of Self-Reproducing Automata*, *op. cit.*, pág. 77.

22. Pattee y Rączaszek-Leonardi, *Laws, Language and Life*, *op. cit.*, pág. 13.

23. Rudolf K. Allemann y Nigel S. Scrutton (comps.), *Quantum Tunneling in Enzyme Catalyzed Reactions*, Cambridge, *Royal Society of Chemistry Publishing*, 2009.

24. Indranil Chakrabarty y Prashant Prashant, «Non Existence of Quantum Mechanical Self Replicating Machine», arXiv: quant- ph/0510221 v6, 2007.

25. Niels Bohr, «The Quantum Postulate and the Recent Development of Atomic Theory», *Nature*, 121, 1928, pág. 580.

26. Pattee, «The Complementarity Principle», *op. cit.*, pág. 144.

27. *Ibid.*, pág. 149.

28. *Ibid.*, pág. 153.

29. Feynman, «Robb Memorial Lecture», *op. cit.*

30. James, *Principles of Psychology*, *op. cit.*, pág. 117.

31. Pattee y Rączaszek-Leonardi, *Laws, Language and Life*, *op. cit.*, pág. 28.

32. *Ibid.*, pág. 11.

1. James, *Principles of Psychology*, *op. cit.*, pág. 149.

2. Matthew E. Roser y otros, «Dissociating Processes Supporting Causal Perception and Causal Inference in the Brain», *Neuropsychology*, 19, 2005, pág. 591.

3. Sherrington, *Man on His Nature*, op. cit. op. cit., pág. 275.

4. Niels Kaj Jerne, «Antibodies and Learning: Selection versus Instruction», en Gardner C. Quarton, Theodore Melnechuk y Francis O. Schmitt (comps.), *The Neurosciences: A Study Program*, Nueva York, Rockefeller University Press, 1967, págs. 200-205.

5. Alan M. Leslie y Stephanie Keeble, «Do Six-Month-Old Infants Perceive Causality?», *Cognition*, 25, 1987, págs. 265-288.

6. Thomas Nagel, «What Is It Like to Be a Bat?», *Philosophical Review*, 83, 1974, págs. 435-450.

7. Liane Young y Rebecca Saxe, «Innocent Intentions: A Correlation between Forgiveness for Accidental Harm and Neural Activity», *Neuropsychology*, 47, 2009, págs. 2.065-2.072.

8. Michael B. Miller y otros, «Abnormal Moral Reasoning in Complete and Partial Callosotomy Patients», *Neuropsychologia*, 48, 2010, págs. 2.215-2.220.

9. Documental sobre Neil Young, parte 2, consultado el 27 de agosto de 2016 en <<https://www.youtube.com/watch?v=Lslh6hG9EVQ>>.

10. Steven Pinker, *Cómo funciona la mente* (2001); cita: *How the Mind Works*, Nueva York, W. W. Norton, 1997, pág. 133.

11. Jaak Panksepp y Lucy Biven, *The Archaeology of Mind: Neuroevolutionary Origins of Human Emotions*, Nueva York, W. W. Norton, 2012.

12. Jaack Panksepp, «The Periconscious Substrates of Consciousness: Affective States and the Evolutionary Origins of the SELF», *Journal of Consciousness Studies*, 5, 1998, págs. 566-582.

13. Andrew R. Barron y Colin Klein, «What Insects Can Tell Us about the Origins of Consciousness», PNAS, 113, 2016, págs. 4.900-4.908.

14. Nicholas J. Strausfeld y Frank Hirth, «Deep Homology of Arthropod Central Complex and Vertebrate Basal Ganglia», *Science*, 340, 2013, págs. 157-161.

15. Robert B. Barlow Jr. y Anthony J. Fraioli, «Inhibition in the Limulus Lateral Eye In Situ», *Journal of General Physiology*, 71, 1978, págs. 699-720.

16. Shreesh P. Mysore y Eric I. Knudsen, «A Shared Inhibitory Circuit for Both Exogenous and Endogenous Control of Stimulus Selection», *Nature Neuroscience*, 16, 2013, págs. 473-478.

17. Diane M. Beck y Sabine Kastner, «Top-down and Bottom-up Mechanisms in Biasing Competition in the Human Brain», *Vision Research*, 49, 2009, págs. 1.154-1.165.

18. Steven D. Wiederman y David C. O'Carroll, «Selective Attention in an Insect Visual Neuron», *Current Biology*, 23, 2013, págs. 156-161.

19. Paul Buckley y F. David Peat, *Glimpsing Reality: Ideas in Physics and the Link to Biology* (ed. rev.), Nueva York, Routledge, 2009, pág. 134.

20. Bisiach y Luzzatti, «Unilateral Neglect of Representational Space», *op. cit.*

21. Denise Barbut y Michael S. Gazzaniga, «Disturbances in Conceptual Space Involving Language and Speech», *Brain*, 110, 1987, págs. 1.487-1.496.

22. Gary Taubes, *Good Calories, Bad Calories: Fats, Carbs, and the Controversial Science of Diet and Health*, Nueva York, Anchor Books, 2007.

1. Howard Hunt Pattee, «Can Life Explain Quantum Mechanics?», en Ted Bastin (comp.), *Quantum Theory and Beyond: Essays and Discussions Arising from a Colloquium*, Cambridge, Cambridge University Press, 1971, págs. 307-319.

2. Rosen, *Life Itself*, *op. cit.*, pág. 13.

3. *Ibid.*, pág. 111.

4. *Ibid.*, pág. 112.

5. William James, «What Is an Instinct?», *Scribner's Magazine*, 1, 1887, pág. 355.

6. *Ibid.*, pág. 356.

7. Joshua T. Vogelstein y otros, «Discovery of Brainwide Neural-Behavioral Maps via Multiscale Unsupervised Structure Learning», *Science*, 344, 2014, págs. 386-392.

8. James, «What Is an Instinct?», *op. cit.*, pág. 359.

El instinto de la conciencia
Michael S. Gazzaniga

No se permite la reproducción total o parcial de este libro, ni su incorporación a un sistema informático, ni su transmisión en cualquier forma o por cualquier medio, sea éste electrónico, mecánico, por fotocopia, por grabación u otros métodos, sin el permiso previo y por escrito del editor. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual (Art. 270 y siguientes del Código Penal).

Diríjase a CEDRO (Centro Español de Derechos Reprográficos) si necesita reproducir algún fragmento de esta obra. Puede contactar con CEDRO a través de la web www.conlicencia.com o por teléfono en el 91 702 19 70 / 93 272 04 47.

Título original: *The Consciousness Instinct*
Publicado originalmente en inglés por Farrar, Straus and Giroux

© del diseño de la portada, Matt Chase
© adaptación del diseño de la portada, Planeta Arte & Diseño
© de la ilustración de la portada, Matt Chase

© Michael S. Gazzaniga, 2018

© de la traducción, Francisco J. Ramos Mena, 2019

© de todas las ediciones en castellano,
Editorial Planeta, S. A., 2019
Paidós es un sello editorial de Editorial Planeta, S. A.
Av. Diagonal, 662-664, 08034 Barcelona (España)
www.planetadelibros.com

Primera edición en libro electrónico (epub): junio de 2019

ISBN: 978-84-493-3608-9 (epub)

Conversión a libro electrónico: Realización Planeta



EL INSTINTO DE LA CONCIENCIA



MICHAEL S. GAZZANIGA



CÓMO EL CEREBRO CREA LA MENTE

PAIDÓS