

EL MISTERIO DE LA CONCIENCIA

JOHN R. SEARLE

Traducción de Juan Almela



Las siguientes páginas son la primera parte de un ensayo del filósofo John R. Searle sobre uno de los problemas centrales del pensamiento científico de nuestros días: el de las relaciones entre el pensamiento y el cerebro (y, si se quiere, entre la mente y la materia). Searle examina con particular minucia la pertinencia de la analogía entre los procesos computacionales y el funcionamiento de la mente. El ensayo parte de la revisión de algunos libros recientes sobre el tema, de Francis Crick, Gerald Edelman y Roger Penrose. La conclusión de este importante ensayo, que publicamos en una versión editada por Vuelta con autorización de The New York Review of Books, aparecerá en nuestro número de marzo.

1

El problema más importante en las ciencias biológicas es uno que, hasta hace muy poco, muchos científicos no consideraban en absoluto como un tema adecuado para la investigación científica. Es éste: ¿cómo exactamente causan la conciencia los procesos neurobiológicos del cerebro? La enorme variedad de estímulos que nos afectan —por ejemplo al probar un vino, mirar el cielo, oler una rosa, escuchar un concierto— desencadenan secuencias de procesos neurobiológicos que a fin de cuentas causan estados subjetivos, internos, coherentes, bien ordenados, unificados, de conciencia y sensación. Ahora bien, ¿qué ocurre exactamente entre el acoso de los estímulos a nuestros receptores y la experiencia de la conciencia, y cómo exactamente los procesos intermedios causan los estados conscientes?

Por lo demás, el problema no se restringe a los casos perceptivos que he mencionado, sino que incluye las experiencias de acciones voluntarias, así como procesos internos como preocuparse por los impuestos o tratar de recordar el número telefónico de la suegra. Es un hecho pasmoso que en la vida consciente todo, desde sentir dolores, cosquillas y comezónes hasta (escoja lo que usted prefiera) padecer la angustia del hombre postindustrial bajo el capitalismo tardío o experimentar la emoción de esquiar, todo es causado por procesos cerebrales. Hasta donde sabemos, los procesos pertinentes

acontecen en los microniveles de las sinapsis, neuronas, columnas neuronales y agregados celulares. Toda nuestra vida consciente es causada por estos procesos de bajo nivel, pero apenas tenemos la más vaga idea de cómo funciona todo ello.

Bueno, se preguntará, ¿por qué no se ocupan de ello los especialistas adecuados y averiguan cómo funciona? ¿por qué ha de ser más difícil que hallar las causas del cáncer? Ahora bien, hay cierto número de caracteres especiales que vuelven aún más difíciles de resolver los problemas suscitados por las ciencias del cerebro. Algunas de las dificultades son prácticas: por lo que hoy se calcula, el cerebro humano tiene más de cien mil millones de neuronas, cada una de las cuales se conecta sinápticamente con otras, que fluctúan en número entre pocos centenares y muchas decenas de miles. Toda esta estructura enormemente compleja está encajonada en un espacio menor que un balón de fútbol. Además es difícil trabajar con los microelementos cerebrales sin dañarlos o matar al organismo. Por añadidura, aparte de las dificultades prácticas hay varios obstáculos y confusiones de orden filosófico y teórico que toman difícil plantear y responder las preguntas debidas. Por ejemplo, el modo, tan de sentido común, como he formulado la cuestión (“¿cómo los procesos cerebrales causan la conciencia?”), lleva ya una carga filosófica. Muchos filósofos, y hasta algunos científicos, piensan que la relación no puede ser causal, ya que una relación causal entre cerebro y conciencia les parece implicar alguna versión del dualismo cerebro-conciencia que desean rechazar por otros motivos.

Desde la época de los antiguos griegos hasta los últimos modelos computacionales de la cognición, el tema entero de la conciencia y de su relación con el cerebro ha representado bastante lío, y al menos algunos de los errores de la historia del tema se repiten en los libros que aquí vamos a reseñar. Antes de discutir la labor más reciente, quiero preparar el escenario poniendo en claro algunos puntos y corrigiendo algunos de los que me parecen peores errores históricos.

Es posible despachar en seguida una cuestión. Existe un problema supuestamente arduo pero que a mí no

me parece muy grave: el problema de definir la conciencia. Se supone que es pavorosamente difícil definir este término. Pero si diferenciamos entre definiciones analíticas, que aspiran a analizar la esencia subyacente de un fenómeno, y definiciones de sentido común, que nada más identifican de qué hablamos, no me parece nada difícil dar una definición de la palabra según el sentido común: "conciencia" refiere a esos estados de sentir y captar que se inician, típicamente, cuando despertamos de un dormir sin sueños y que continúan hasta que volvemos a dormirnos o caemos en coma o morimos o de algún otro modo quedamos "inconscientes". Los sueños son una forma de conciencia, si bien, por supuesto, muy distinta de estados de plena vigilia. La conciencia así definida funciona y deja de funcionar. Según esta definición, un sistema es consciente o no lo es, pero dentro del terreno de la conciencia se da una gama de estados de intensidad que van desde el adormilamiento hasta la plena atención. La conciencia así definida es un fenómeno cualitativo, interior, de primera persona. Los seres humanos y los animales superiores están evidentemente conscientes, pero ignoramos hasta dónde se extiende la conciencia al descender la escala filogenética.

¿Están conscientes, por ejemplo, las pulgas? En el estado presente del conocimiento neurobiológico probablemente no es provechoso cuidarse de tales preguntas. No sabemos biología suficiente para saber dónde está la frontera.

El primer problema serio deriva de la historia intelectual. En el siglo XVII Descartes y Galileo establecieron una rotunda distinción entre la realidad física descrita por la ciencia y la realidad mental del alma, que consideraban fuera del alcance de la investigación científica. Este dualismo de mente consciente y materia inconsciente fue útil en la investigación científica de entonces, tanto porque ayudaba a que los científicos se quitasen de encima a las autoridades religiosas, como porque el mundo físico era matemáticamente abordable de un modo que la mente no parecía compartir. Sólo que este dualismo se ha convertido en el siglo XX en un obstáculo, ya que parece situar la conciencia y otros fenómenos mentales fuera del mundo físico ordinario y, así, fuera del dominio de la ciencia natural. A mi modo de ver, tenemos que abandonar el dualismo y partir del supuesto de que la conciencia es un fenómeno biológico ordinario comparable al crecimiento, la digestión o la secreción de la bilis. Pero muchas personas que laboran en las ciencias se mantienen dualistas y no creen que podamos ofrecer una explicación causal de la conciencia que la muestre como parte de la realidad biológica ordinaria. Quizá el más famoso de este grupo sea el neurobiólogo Sir John Eccles, premio Nobel, quien cree que Dios liga el alma al feto nonato a la edad de unas tres semanas.

De los autores de quienes aquí nos ocupamos, Roger Penrose es un dualista en el sentido de opinar que no vivimos en un mundo unificado, sino de que, más bien, hay un mundo mental separado que "se funda" en el mundo físico. De hecho, cree que vivimos en tres mundos, un mundo físico, un mundo mental y un mundo de objetos abstractos, como los números y otras entidades matemáticas. Luego me ocuparé más de esto.

Pero incluso si tratamos la conciencia como un fenómeno biológico y, de este modo, como parte del mundo físico ordinario —según insisto en que debiéramos hacerlo—, quedan aún muchos otros errores por evitar. Uno lo acabo de mencionar: si los procesos cerebrales causan la conciencia, entonces a mucha gente le parece que debe haber dos cosas diferentes, procesos cerebrales como causas y estados conscientes como efectos, lo cual parece implicar dualismo. Esta segunda equivocación deriva en parte de un concepto viciado de la causación. En nuestras teorías oficiales de la causación, lo típico es suponer que todas las relaciones causales deben darse entre acontecimientos discretos ordenados sucesivamente en el tiempo. Por ejemplo, el tiroto causó la muerte de la víctima.

Claro es que muchas relaciones de causa y efecto son así, pero de ninguna manera todas. Mfrese alrededor los objetos cercanos, y piénsese en la explicación causal del hecho de que la mesa ejerza presión sobre la alfombra. Esto lo explica la fuerza de gravedad, pero la gravedad no es un acontecimiento. O piénsese en la solidez de la mesa. Se explica causalmente por el comportamiento de las moléculas de que la mesa se compone. Pero la solidez de la mesa no es un suceso adicional, sino sencillamente una característica de la mesa. Semejantes ejemplos de causación sin acontecimiento nos ofrecen modelos apropiados para comprender la relación entre mi presente estado de conciencia y los procesos neurobiológicos que lo causan. Procesos cerebrales a bajo nivel causan mi presente estado de conciencia, pero dicho estado no es una entidad separada de mi cerebro; es, más bien, una característica de mi cerebro en este momento. De paso, este análisis —que los procesos cerebrales *causan* la conciencia pero que esta misma es una *característica* del cerebro— nos proporciona una solución del problema tradicional mente-cuerpo, solución que evita tanto el dualismo como el materialismo, cuando menos tal como tradicionalmente se conciben.

Como tercera dificultad de nuestra presente situación intelectual mencionemos que no tenemos una idea nada clara de cómo los procesos cerebrales, que son fenómenos objetivos, públicamente observables, pudieran causar algo tan peculiar como estados cualitativos internos de captar o sentir; estados, en algún sentido, "privados" para el poseedor del estado. Mi dolor tiene cierto matiz cualitativo y me es accesible de

un modo que no le es accesible a usted. Ahora bien, ¿cómo *podieran* estos fenómenos cualitativos, subjetivos, privados, ser causados por procesos físicos ordinarios, tales como disparos electroquímicos en las sinapsis de neuronas? Cada tipo de estado consciente tiene un matiz cualitativo especial, y no concordamos acerca de cómo insertar estas impresiones subjetivas en nuestra visión total del mundo como consistente en realidad objetiva. Tales estados y acontecimientos son llamados a veces *qualia* y el problema de dar razón de ellos dentro de nuestra visión total del mundo se denomina el problema de los *qualia*. Entre las diferencias interesantes entre las exposiciones de los autores que ahora reseñamos, están sus múltiples y divergentes maneras de vérselas —o de no lograrlo— con el problema de los *qualia*.

La cuarta dificultad es peculiar de nuestro clima intelectual precisamente ahora, y es el apremio a tomar demasiado literalmente la metáfora de la mente como computadora. Muchas personas siguen creyendo que el cerebro es una computadora digital y que la mente consciente es un programa de computadora, si bien este modo de ver, afortunadamente, ya está menos difundido que hace diez años. Desde este punto de vista, la mente es al cerebro como el *software* es al *hardware*. Hay diferentes versiones de la teoría computacional de la mente. La más estricta es la que acabo de enunciar: la mente no es sino un programa de computadora. No hay más. A esta actitud la llamo inteligencia artificial fuerte (IA-fuerte, para abreviar), a fin de distinguirla de la opinión de que la computadora es una herramienta útil para crear simulaciones de la mente, igual que de casi cualquier cosa que podamos describir con precisión, tales como las pautas climáticas o el curso del dinero en la economía. Este modo más cauto de ver lo llamo IA-débil.

La IA-fuerte puede refutarse rápidamente, según lo hice hace más de una década ("The Myth of the Computer", *The New York Review of Books*, 29 de abril, 1982; y "Minds, Brains and Programs", *Behavioral and Brain Sciences*, 1980, vol. 3, pp. 417-457). Una computadora es por definición un artefacto que maneja símbolos formales. Se acostumbra describirlos como ceros y unos, aunque cualquier otro símbolo serviría lo mismo. La computación, así definida, es un conjunto puramente sintáctico de operaciones. Pero sabemos por experiencia propia que en la mente se da algo más que manejo de símbolos formales; las mentes tienen contenidos. Por ejemplo, si pensamos en español, las palabras españolas que cruzan nuestras mentes no son meros símbolos formales sin interpretar; sabemos, por el contrario, lo que significan. Para nosotros las palabras tienen significado, semántica. La mente no podría ser nada más un programa de computadora porque los símbolos formales de dicho programa, por sí mismos, no bastan para

garantizar la presencia del contenido semántico que se da en las mentes reales.

He ilustrado este punto con un sencillo experimento mental. Imagínesse que seguimos los pasos de un programa a fin de contestar preguntas en un lenguaje que no entendemos. Yo no entiendo el chino, así que me imagino encerrado en un cuarto entre muchas cajas de símbolos chinos (la base de datos); recibo pequeños manojos de símbolos chinos (preguntas en chino), y consulto en un libro de reglas (el programa) qué se espera que haga yo. Realizo ciertas operaciones con los símbolos de acuerdo con las reglas (esto es, sigo los pasos del programa) y devuelvo pequeños manojos de símbolos a quienes están afuera del cuarto (respuestas a las preguntas). Soy la computadora implementando un programa para contestar preguntas en chino, pero no entiendo, con todo, ni una palabra de chino. Y aquí está a lo que iba: *si no entiendo el chino sobre la pura base de implementar un programa de computadora para entender el chino, entonces tampoco ninguna otra computadora digital lo conseguirá sobre esa pura base, puesto que ninguna computadora digital posee nada que yo no posea.*

Este argumento es tan sencillo y decisivo que me embaraza tener que repetirlo, pero en los años transcurridos desde que lo publiqué por vez primera debe de haber aparecido más de un centenar de ataques. El "argumento del cuarto chino" —como ha pasado a llamarse— tiene una simple estructura en tres pasos:

1. Los programas son sintácticos.
2. Las mentes tienen semántica.

3. La sintaxis no es lo mismo que la semántica, ni es por sí misma suficiente.

Por lo tanto los programas no son mentes, q.e.d.

A fin de refutar el argumento habría que mostrar que una de estas premisas es falsa, lo cual no es nada probable.

Ahora me parece que este argumento, si acaso, concede demasiado a la IA-fuerte, pues concede que la teoría era cuando menos falsa. Ahora pienso que es incoherente, y he aquí por qué. Preguntémosnos cuál hecho acerca de la máquina donde estoy escribiendo esto hace sintácticas o simbólicas sus operaciones. Por lo que a su física respecta, no es sino un circuito electrónico muy complejo. El hecho que hace simbólicos estos pulsos eléctricos es la misma clase de hecho que torna símbolos las marcas de tinta de las páginas de un libro: hemos diseñado, programado, impreso y manufacturado estos sistemas de suerte que podamos tratar y usar semejantes cosas como símbolos. En una palabra, la sintaxis no es intrínseca a la física del sistema sino que está en el ojo del que mira. Salvo por los pocos casos de agentes conscientes que de hecho realizan una computación —sumar 2+2 para obtener 4, digamos—, la computación no es un proceso intrínseco de la naturaleza, como la digestión o la fotosíntesis, sino que sólo existe

relativamente a algún agente que asigna una interpretación computacional a la física. La consecuencia es que la computación no es intrínseca a la naturaleza sino que es relativa al observador o usuario.

La consecuencia para nuestra discusión presente es que la pregunta de si el cerebro será una computadora digital carece de sentido claro. Si pregunta: "¿es el cerebro intrínsecamente una computadora digital?" la respuesta, trivial, es que no, porque nada es intrínsecamente una computadora digital; algo lo es solamente en relación con la atribución de una interpretación computacional. Si pregunta "¿puede asignarse una interpretación computacional al cerebro?" la respuesta, trivial, es que sí, ya que puede asignarse una interpretación computacional a cualquier cosa. Por ejemplo, la ventana que tengo delante es una computadora muy sencilla. Ventana abierta = 1, ventana cerrada = 0. Jamás podrían descubrirse procesos computacionales en la naturaleza independientemente de interpretación humana, porque cualquier proceso físico que pudiera encontrarse no es computacional sino relativamente a alguna interpretación. Este es un punto evidente y debí haberlo notado hace mucho.

La consecuencia es que la IA-fuerte, que se enorgullece de su "materialismo" y de la visión del cerebro como una máquina, no es lo bastante materialista. El cerebro es en efecto una máquina, una máquina orgánica, y sus procesos, tales como los disparos neuronales, son procesos de máquina orgánica. Pero la computación no es un proceso de máquina, como el disparo neuronal o la combustión interna; la computación es, más bien, un proceso matemático abstracto que existe sólo en relación a observadores e intérpretes conscientes. Observadores como nosotros han hallado maneras de implementar la computación en máquinas eléctricas con base de silicio, pero esto no convierte la computación en algo eléctrico o químico.

Este argumento difiere del del cuarto chino pero es más profundo. El argumento del cuarto chino mostró que la semántica no es intrínseca a la sintaxis; éste muestra que la sintaxis no es intrínseca a la física.

Rechazo la IA-fuerte pero acepto la IA-débil. De los autores que estoy reseñando, Roger Penrose rechaza incluso la IA-débil. Piensa que la mente no puede siquiera ser simulada con una computadora. Gerald Edelman acepta el argumento del cuarto chino contra la IA-fuerte y presenta algunos otros argumentos de su cosecha, pero acepta la IA-débil; hasta aprovecha con vigor modelos de computadora en su investigación del cerebro, según veremos.

Resumiendo la posición general, pues, en torno a cómo puede proceder la investigación del cerebro para responder las cuestiones que nos acosan: el cerebro es un órgano como cualquier otro; es una máquina orgánica. La conciencia es causada por procesos neuronales

de bajo nivel en el cerebro, y es ella misma una característica del cerebro.

Dado que es una característica procedente de ciertas actividades neuronales, podemos concebirla como una "propiedad emergente" del cerebro. Las computadoras desempeñan en el estudio del cerebro el mismo papel que en cualquier otra disciplina. Son aparatos inmensamente provechosos para simular procesos cerebrales. Pero la simulación de estados mentales no es un estado mental en grado mayor que la simulación de una explosión es en sí misma una explosión.

2

He afirmado que hasta hace poco había entre los científicos resistencia a enfrentarse al problema de la conciencia. Ahora todo esto ha cambiado y ha aparecido una serie de libros debidos a biólogos, matemáticos y físicos, a más de filósofos. De los libros reseñados, el que intenta presentar del modo más sencillo y directo lo que sabemos acerca de cómo funciona el cerebro es de Crick, *The Astonishing Hypothesis: The Scientific Search for the Soul* ("la hipótesis asombrosa: la búsqueda científica del alma"). La asombrosa hipótesis es que el libro se funda es

que 'tú', tus alegrías y tus penas, tus recuerdos y tus ambiciones, tu sentido de identidad personal y libre albedrío, no son de hecho más que el comportamiento de un gran agregado de células nerviosas y las moléculas que se les asocian.

He visto reseñas del libro de Crick que protestaban diciendo que nada tiene de asombroso, hoy por hoy, que le digan a uno que lo que ocurre dentro de nuestros cráneos es responsable de toda nuestra vida mental, y que cualquiera que disfrute de un mínimo de formación científica aceptará la hipótesis de Crick como una especie de trivialidad. No creo que semejante protesta sea justa. El asombro de Crick tiene dos partes. La primera es que toda nuestra vida mental tiene una existencia material en el cerebro —lo cual realmente no es muy asombroso—, pero también, lo cual es más interesante, que los mecanismos específicos del cerebro responsables de nuestra vida mental son neuronas y moléculas que se les asocian, tales como las moléculas de neurotransmisores. Por mi parte, siempre me asombra la especificidad de los sistemas biológicos, y en el caso del cerebro la especificidad asume una forma que no hubiera podido predecirse, sin más, conociendo lo que hace. Puestos a proyectar una máquina orgánica para bombear sangre, podría irse a parar a algo semejante a un corazón, pero, tratándose de proyectar una máquina productora de conciencia, ¿a quién se le iban a ocurrir cien mil millones de neuronas?

Crick no es claro al diferenciar las explicaciones

causales de la conciencia, de las eliminaciones reduccionistas de la conciencia. El pasaje que acabo de citar da la impresión de que negara que tuviésemos experiencias conscientes además de tener disparos neuronales. Pero pienso que una lectura cuidadosa del libro muestra que lo que quiere decir es algo parecido a la tesis que propuse antes: todas nuestras experiencias conscientes son explicadas por el comportamiento de neuronas y son ellas mismas propiedades emergentes del sistema de neuronas. (Una propiedad emergente de un sistema es la que es explicada por la conducta de los elementos del sistema, pero no es propiedad de ningunos elementos individuales ni puede explicarse sencillamente como suma de las propiedades de esos elementos. La liquidez del agua es un buen ejemplo: el comportamiento de las moléculas de H₂O explica la liquidez pero cada una de las moléculas no es líquida.)

El hecho de que la parte explicativa de lo que plantea Crick sea hoy en día la ortodoxia neurobiológica ordinaria y asombre a pocos de quienes se inclinan a leer el libro, no debe impedirnos convenir en que es asombroso cómo el cerebro hace tanto con mecanismo tan limitado. Por lo demás, no todos quienes laboran en este campo concuerdan en que la neurona sea el elemento funcional esencial. Penrose cree que ya las neuronas son demasiado grandes y aspira a dar razón de la conciencia en un nivel de fenómenos mecánico-cuánticos mucho menores. Edelman encuentra las neuronas demasiado pequeñas para la mayoría de las funciones y considera "grupos neuronales" como los elementos funcionales.

Crick emplea la percepción visual como cuña para intentar irrumpir en el problema de la conciencia. Opino que es una buena elección, así fuera sólo por haberse trabajado tanto acerca de la anatomía y fisiología de la vista. Pero un problema con esto es que el funcionamiento del sistema visual es pasmosamente complicado. El simple acto de reconocer la cara de un amigo entre una multitud requiere, en cuanto a procesamiento, mucho más de lo que alcanzamos a comprender al presente. Yo conjeturo —y no pasa de ser una conjetura— que cuando por fin irrumpamos en la comprensión de cómo los disparos neuronales causan conciencia o sensación, será seguramente gracias a entender un sistema sencillo.

Esto sin embargo no disminuye el interés, la verdadera fascinación, con que Crick nos guía por las capas de células de la retina hasta el núcleo geniculado lateral, y de ahí a la corteza visual y por último a las diferentes áreas de la corteza.

¿Cómo funciona la neurona? Una neurona es una célula como cualquier otra, con su membrana celular y su núcleo central. Las neuronas difieren, sin embargo, de otras células, de ciertas maneras tanto anatómicas como fisiológicas. Hay muchos tipos diferentes de neu-

ronas, pero la variedad corriente típica lleva en un lado una protuberancia alargada, filamentosa —llamada axón— y del otro lado un puñado de fibras espinosas, puntiagudas, algo más cortas, denominadas dendritas. Cada neurona recibe señales por sus dendritas, las procesa en su cuerpo celular o soma y luego dispara una señal por su axón hacia las siguientes neuronas en línea. La neurona dispara enviando un impulso eléctrico por el axón. El axón de una neurona, con todo, no está conectado directamente con las dendritas de otras, sino que el punto donde la señal es transmitida de una célula a la siguiente es una pequeña brecha que se llama sinapsis. La sinapsis típica consiste en un abultamiento del axón, un "botón", que se proyecta más o menos con forma de hongo y por lo general se acopla junto a una prolongación parecida a una espina en la superficie de la dendrita. El área que hay entre el botón y la superficie dendrítica postsináptica es la brecha sináptica y cuando la neurona dispara, la señal es transmitida a través de ese espacio.

La señal no es transmitida mediante una conexión eléctrica directa entre botón y superficie dendrítica, sino mediante la liberación de pequeñas cantidades de sustancias llamadas neurotransmisores. Cuando la señal eléctrica llega desde el cuerpo celular por el axón, hasta el extremo de un botón, provoca la liberación de líquidos neurotransmisores en la brecha sináptica. Establecen contacto con receptores situados del lado dendrítico postsináptico. Se abren compuertas y pasan iones cargados hacia el lado dendrítico o desde él, alterando con ello el potencial eléctrico de la dendrita. La sucesión es, pues, ésta: una señal eléctrica del lado del axón, seguida de transmisión química por la brecha sináptica, seguida de una señal eléctrica del lado dendrítico. La célula recibe multitud de señales por sus dendritas, las suma en su soma y, de acuerdo con el resultado, ajusta su ritmo de disparo hacia las siguientes células de la línea.

Las neuronas reciben señales tanto excitadoras —o sea que tienden a incrementar su ritmo de disparo— como inhibitorias, que tienden a disminuir dicho ritmo. Cosa bastante singular, aunque cada neurona reciba señales tanto excitadoras como inhibitorias, cada neurona sólo envía señales de uno de los tipos. Hasta donde sabemos, con escasas excepciones, una neurona es excitadora o es inhibitoria.

Ahora bien, lo asombroso es esto: por lo que atañe a nuestra vida mental, lo que acabo de relatar acerca de las neuronas constituye toda la base causal de nuestra vida consciente. He descartado varios detalles acerca de canales iónicos, receptores y diferentes tipos de neurotransmisores, pero nuestra vida mental entera es causada por el comportamiento de neuronas, y todo lo que hacen es incrementar o reducir sus tasas de disparo. Cuando, por ejemplo, almacenamos recuerdos, se diría

que debemos hacerlo, de algún modo, en las conexiones sinápticas entre neuronas. Crick logra notablemente bien no solamente resumir la exposición simple del funcionamiento cerebral sino también procurar su integración con el trabajo en muchos terrenos afines, incluyendo la psicología, la ciencia cognitiva y el uso de computadoras para establecer modelos de redes neuronales, entre otras cosas.

Crick es generalmente hostil hacia los filósofos y la filosofía, pero el despreciar la filosofía hace cometer errores filosóficos. La mayoría de éstos no perjudican seriamente su argumentación principal, aunque sean molestos y superfluos. Mencionaré tres errores filosóficos que sí considero graves.

En primer lugar, Crick entiende mal el problema de los *qualia*. Cree que es ante todo un problema acerca de cómo una persona tiene conocimiento de los *qualia* de otra persona. "El problema surge del hecho de que la rojez del rojo que tan vívidamente percibo no pueda ser comunicada con precisión a otro ser humano." Sólo que éste no es el problema o, mejor dicho, sólo es una pequeña parte de él. Aun a propósito de un sistema cuyos *qualia* conozca yo casi a la perfección —yo mismo, por ejemplo—, el problema de los *qualia* es serio. Es éste: ¿cómo es posible que disparos neuronales físicos, objetivos, cuantitativamente describibles, provoquen experiencias subjetivas, cualitativas, privadas? ¿Cómo, por plantearlo ingenuamente, salva el cerebro el monte que separa electroquímica y sensación? He aquí la parte dura del problema mente-cuerpo, que persiste después de advertir que la conciencia debe ser causada por procesos cerebrales y es, ella misma, una característica del cerebro.

Además, el problema de los *qualia* no es tan sólo un aspecto del problema de la conciencia; es el problema de la conciencia. Puede hablarse de varios otros rasgos de ésta —por ejemplo, de la capacidad del sistema visual para discriminar colores—, pero en la medida en que se esté hablando de discriminación consciente, se está hablando de *qualia*. Por esta razón opino que el término es equívoco, pues sugiere que los *qualia* de un estado de conciencia podrían desprenderse del resto de la conciencia y dejarse a un lado, como si se pudiera hablar del resto del problema de la conciencia, ignorando el sentir subjetivo, cualitativo de la conciencia. Sólo que no pueden dejarse de lado los *qualia*, pues si se hace esto ya no queda conciencia.

Segundo, Crick es inconsistente en su exposición de la reducción de la conciencia a disparos neuronales. Lo que dice suena a reduccionismo pero su exposición no es de naturaleza reduccionista en absoluto. O, más bien, hay dos sentidos de "reducción" que no consigue distinguir. En un sentido, la reducción es eliminativa. Nos quitamos de encima el fenómeno reducido mostrando que realmente es otra cosa. Las puestas de sol

sirven bien de ejemplo. En realidad el sol no se hunde tras la montaña; el aspecto del sol poniéndose es una ilusión explicada enteramente por la rotación de la tierra sobre su eje con relación al sol. Pero en otro sentido de reducción explicamos un fenómeno pero no nos lo quitamos de encima. Así, la solidez de un objeto es explicada enteramente por la conducta molecular, lo cual tampoco revela que ningún objeto sea verdaderamente sólido, ni que no exista distinción entre, digamos, solidez y liquidez. Ahora bien, Crick habla como aspirando a una reducción eliminadora de la conciencia, pero de hecho su libro se orienta hacia una explicación causal. No intenta mostrar que la conciencia no exista realmente. Abandona el juego cuando dice, correctísimamente en mi opinión, que la conciencia es una propiedad emergente del cerebro. Las sensaciones complejas, dice, son propiedades emergentes que "surgen en el cerebro merced a las interacciones de sus muchas partes".

Lo desconcertante es que Crick predique reduccionismo eliminador mientras practica emergentismo causal. El argumento ordinario en filosofía, contra una reducción eliminadora de la conciencia es que, por mucho que poseyéramos una perfecta ciencia neurobiológica; seguiría habiendo dos cosas distintas, la pauta neurobiológica de disparos neuronales y la sensación de dolor, por ejemplo.

Los disparos neuronales causan la sensación pero no son lo mismo que ésta. En diferentes versiones este argumento es adelantado, entre otros filósofos, por Thomas Nagel, Saul Kripke, Frank Jackson y el presente autor. Crick, erradamente, apoya un argumento contra este antirreduccionismo, debido a Paul y Patricia Churchland. Es un mal argumento y bien puedo decir por qué me lo parece. Piensan que el planteamiento filosófico ordinario se basa en la siguiente argumentación, evidentemente errada:

1. Pedro sabe que su sensación es un dolor.
2. Pedro no sabe que su sensación es una pauta de disparos neuronales.
3. Por lo tanto la sensación de Pedro no es una pauta de disparos neuronales.

(Han propuesto este argumento en varios lugares, el más reciente *The Engine of Reason, the Seat of the Soul*, por Paul M. Churchland, MIT Press, 1995, pp. 195-208.)

Los Churchland señalan que este argumento es falaz, pero se equivocan pensando que es el que utilizan sus adversarios. La argumentación que atacan es "epistémica", trata del conocimiento. Pero la que atañe a la irreductibilidad de la conciencia no lo es, trata de cómo las cosas son en el mundo. Trata de ontología. Hay diferentes modos de exponerla pero en todos es el mismo el punto principal: la pura sensación cualitativa de dolor es un rasgo cerebral muy diferente de la pauta de

disparos neuronales que causan dicho dolor. De modo que se logra una reducción causal de dolor a disparos neuronales pero no una reducción ontológica. Esto es, puede darse una explicación causal completa de por qué sentimos dolor, sólo que eso no muestra que realmente no existan los dolores.

En tercer lugar, Crick no es claro acerca de la estructura lógica de la explicación que ofrece, y aun leyéndolo con la mayor simpatía parece inconsistente. Hasta aquí lo he estado interpretando como en busca de una explicación causal de la conciencia visual, y lo que dice sobre "mecanismo" y "explicación" apoya tal interpretación. Pero nunca dice a las claras que está ofreciendo una explicación causal de cómo los procesos cerebrales causan la conciencia visual. El modo de hablar que prefiere consiste en afirmar que va en pos de los "correlatos neurales" de la conciencia. Sólo que, según sus propios términos, "correlatos neurales" no puede ser la expresión acertada. Ante todo, una correlación es una relación entre dos cosas diferentes, salvo que una relación entre dos cosas es inconsistente con la tesis reduccionista eliminadora que Crick cree estar asumiendo. Para el modo de ver reduccionista eliminador sólo debiera haber una cosa, disparos neuronales. Segundo y más importante: aunque se elimine el error reduccionista, las correlaciones por sí mismas no explicarían nada. Piénsese en la apariencia del rayo y el ruido del trueno. Apariencia y ruido están perfectamente correlacionados, si bien en ausencia de una teoría causal no se obtiene una explicación.

Crick, por añadidura, tampoco es claro a propósito de la relación entre las experiencias visuales y los objetos del mundo de los cuales las experiencias son. Unas veces dice que la experiencia visual es una "descripción simbólica" o "interpretación simbólica" del mundo. Otras veces dice que los procesos neuronales "representan" objetos del mundo. Hasta se ve conducido a negar que tengamos conciencia perceptiva directa de los objetos del mundo, y para alcanzar dicha conclusión aprovecha un mal argumento que viene directamente de la filosofía del seiscientos. Sostiene que, como nuestras interpretaciones pueden a veces ser equivocadas, carecemos de conocimiento directo de los objetos del mundo. Este argumento figura tanto en Descartes como en Hume, pero es una falacia. Del hecho de que nuestras experiencias perceptivas estén siempre mediadas por procesos cerebrales (¿de qué otro modo tenerlas?) y el hecho de que se presten típicamente a ilusiones de varias clases, no se sigue que nunca veamos el mundo real sino sólo "descripciones simbólicas" o "interpretaciones simbólicas" del mundo. En el caso ordinario, como cuando miro el reloj, realmente veo el reloj real. No veo una "descripción" ni una "interpretación" suya.

Creo que Crick ha estado mal aconsejado filosóficamente, pero por fortuna es posible eliminar las confu-

siones filosóficas y encontrar, pese a todo, un libro excelente. Lo que a mi parecer a fin de cuentas desea —y que en todo caso deseo yo— es una explicación causal de la conciencia, ya visual o de otra índole. Los fotones reflejados por los objetos atacan las células fotorreceptoras de la retina, lo cual inicia una serie de procesos neuronales (pues la retina es parte del cerebro), que conducen por último, si todo marcha bien, a una experiencia visual que es una percepción del objeto mismo que empezó por reflejar los fotones. Tal es el modo debido de pensar en ello y creo que Crick estaría de acuerdo.

¿Cuál es pues la solución de Crick al problema de la conciencia? Uno de los rasgos más atractivos de su libro es su disposición, que raya en anhelo, a admitir cuán poco sabemos. Pero dado lo que conocemos, emprende algunas especulaciones. A fin de explicar sus especulaciones en torno a la conciencia, necesito decir algo acerca de lo que los neurobiólogos denominan "el problema del enlace". Sabemos que el sistema visual posee células y aun regiones que responden especialmente a caracteres particulares de los objetos, tales como color, forma, movimiento, líneas, ángulos, etc. Pero cuando vemos un objeto obtenemos una experiencia unificada de un solo objeto. ¿Cómo liga el cerebro todos estos estímulos distintos en una experiencia única, unificada, de un objeto? El problema se extiende a los diversos modos de percepción. Cuanto atañe a mis experiencias presentes forma parte de una gran experiencia consciente unificada (que Kant, con su habitual don para las frases pegajosas, llamó "la unidad trascendental de la apercepción").

Crick dice que el problema del enlace es "el problema de cómo estas neuronas se tornan temporalmente activas como una unidad". Sólo que éste no es el problema del enlace; es más bien un camino posible hacia la solución de dicho problema. Por ejemplo, un camino posible hacia la solución del problema del enlace en la visión ha sido propuesto por varios investigadores, en especial Wolf Singer y sus colaboradores, en Frankfurt. Opinan que la solución pudiera consistir en la sincronización del disparo de neuronas especialmente separadas, correspondientes a los diferentes caracteres de un objeto. Las neuronas que responden a forma, color y movimiento, pongamos por caso, disparan sincrónicamente en la gama general de cuarenta disparos por segundo (cuarenta hercios). Crick y su colega Christof Koch llevan esta hipótesis un paso más adelante y sugieren que acaso el disparo neuronal sincronizado dentro de esta gama (cuarenta hercios más o menos, pero fluctuando entre treinta y cinco y hasta setenta y cinco) fuera el "correlato cerebral" de la conciencia visual.

También el tálamo parece desempeñar un papel central en la conciencia, y en particular ésta parece depender de los circuitos que conectan tálamo y corteza,

especialmente las capas corticales cuatro y seis. De ahí que Crick especule que tal vez el disparo sincronizado en la gama de cuarenta hercios en las redes que unen el tálamo con la corteza podría ser la clave para resolver el problema de la conciencia.

Admiro la disposición de Crick a especular, pero la naturaleza de la especulación revela cuánto nos falta todavía. Supóngase que resultase que la conciencia se correlacione invariablemente con tasas de disparo neuronal de cuarenta hercios en circuitos de neuronas que conectan el tálamo con la corteza cerebral. ¿Sería tal cosa una explicación de la conciencia? No. Jamás lo aceptaríamos, sin más, como explicación. Lo consideraríamos un formidable paso adelante, pero seguiríamos queriendo averiguar cómo funcionaba. Equivaldría a saber que los movimientos de un automóvil están "correlacionados con" la oxidación de hidrocarburos en el motor. Falta averiguar los mecanismos mediante los cuales la oxidación de hidrocarburos produce, o sea causa, el movimiento de las ruedas. Aun si las especulaciones de Crick resultasen cien por ciento correctas, seguiría haciéndonos falta conocer por qué mecanismos los correlatos neurales causan las sensaciones conscientes, y estamos muy lejos de saber siquiera qué forma asumiría semejante explicación.

Crick ha escrito un libro bueno y útil. Sabe mucho y lo explica con claridad. La mayoría de mis objeciones apunta a sus pretensiones y supuestos filosóficos, pero al leer el libro es posible prescindir de las partes filosóficas y simplemente aprender acerca de la psicología de la visión y la ciencia del cerebro. Las limitaciones de las partes neurobiológicas del libro son las limitaciones del tema precisamente ahora: ignoramos cómo se empañan la psicología de la visión y la neurofisiología, y tampoco sabemos cómo los procesos cerebrales causan la conciencia, ya sea visual o de otro género.

3

De los libros reseñados, tal vez el más ambicioso sea *Shadows of the Mind* ("sombras de la mente") de Penrose. Es continuación de su anterior *The Emperor's New Mind* ("la nueva mente del emperador") y vuelven a plantearse aquí muchas de las tesis anteriores, con mayores desarrollos y respuestas a objeciones alzadas por el libro previo. La obra se divide en dos partes iguales. En la primera utiliza una variación de la famosa prueba debida a Gödel de la incompletud de sistemas matemáticos, a fin de demostrar que no somos computadoras ni podemos siquiera ser simulados por éstas. No sólo es falsa la IA-fuerte sino asimismo la IA-débil. En la segunda mitad presenta una larga explicación de la mecánica cuántica así como algunas especulaciones acerca de cómo una teoría mecánico-cuántica del cerebro podría explicar la conciencia de una manera que juzga imposi-

ble para la física clásica. Dicho sea de paso, este es el único libro que conozco donde pueden hallarse explicaciones detenidas y claras de dos de los descubrimientos más importantes del siglo, el teorema de incompletud de Gödel y la mecánica cuántica.

Admiro enormemente tanto a Penrose como sus libros. Es impetuoso, entusiasta, valiente y muchas veces original. Conuerdo con él asimismo en sostener que el modelo computacional de la mente anda mal encaminado, si bien por razones que señalé brevemente antes, y no por las de él. Mi principal desacuerdo es que su aplicación de la prueba de Gödel no me parece lograda, y quisiera decir por qué.

La primera versión de este argumento se remonta a un artículo de John Lucas, filósofo de Oxford, publicado a principios de los años sesenta. Lucas sostuvo que Gödel había mostrado que hay enunciados en sistemas matemáticos que no pueden probarse dentro de dichos sistemas, si bien podemos ver que son verdaderos. Se sigue, de acuerdo con Lucas, que nuestro entendimiento rebasa al de cualquier computadora. Una computadora usa solamente algoritmos, es decir conjuntos de reglas precisas que especifican que ha de ser emprendida una sucesión de acciones a fin de resolver un problema o probar una proposición. Pero ningún algoritmo puede probar lo que podemos ver como verdadero. Se sigue que nuestro conocimiento de semejantes verdades no puede ser algorítmico. En vista de que las computadoras emplean algoritmos nada más (un programa es un algoritmo), se sigue que no somos computadoras.

La argumentación de Lucas suscitó muchas objeciones, entre las cuales ésta es obvia: del hecho de que nuestro conocimiento de dichas verdades no nos llegue a través de un algoritmo demostrador de teoremas, no se sigue que no usemos en absoluto algoritmos para alcanzar tales conclusiones. NO todos los algoritmos son demostradores de teoremas. Es posible que estemos aplicando algún procedimiento computacional que no sea demostrador de teoremas.

Penrose revive la argumentación con una hermosa versión del razonamiento de Gödel, elaborada inicialmente por Alan Turing y que suele ser llamada "la prueba de la insolubilidad del problema de la detención".

¿Qué hacemos con el argumento de que no empleamos algoritmos a fin de cerciorarnos de que sabemos? Puesto así, parece contener una falacia y podemos oponerle una variante de lo que le opusimos a Lucas. A partir del hecho de que ningún algoritmo "cognosciblemente correcto" pueda explicar nuestras capacidades matemáticas, no se sigue que no pudiera haber un algoritmo del cual no supiéramos —y acaso ni pudiéramos saber— que era correcto, y que explicase dichas capacidades. Supóngase que seguimos un programa inconscientemente y que es tan largo y complicado que

jamás consiguiéramos captarlo entero de una vez. Entenderíamos cada paso, pero serían tantos millares que nunca llegaríamos a captar de una vez el programa entero. O sea que la argumentación de Penrose descansa en lo que podemos saber y comprender, pero no es un requisito de la ciencia cognitiva computacional que la gente sea capaz de comprender los programas que se supone que está usando para resolver problemas cognocitivos.

Esta objeción fue hecha por Hilary Putnam en su reciente reseña de Penrose, quien respondió con una carta indignada (*New York Times Book Review*, 20 de noviembre de 1994). Desde hace mucho Penrose está al tanto de dicha objeción y la discute en su libro con cierto detenimiento. A decir verdad, considera y responde en detalle unas veinte objeciones a su razonamiento. Dice que la idea de que estamos usando algoritmos que no conocemos o no podemos conocer o entender "no es plausible en absoluto", y presenta varias razones por las que sostiene esto. Con miras a mi discusión, supondré que Penrose acierta en cuanto a la implausibilidad de suponer que estamos empleando algoritmos desconocidos cuando alcanzamos conclusiones matemáticas. Pero aún suponiendo que sea implausible y que estemos muy convencidos de no estar usando algoritmos desconocidos al pensar, esta objeción técnica prepara el camino a dos objeciones filosóficas más graves. Primero, ¿cuál cree Penrose que es el propósito de los modelos computacionales de la mente? Parece pensar que cuando se llega al razonamiento matemático el propósito es obtener programas que logren hacer lo que los seres humanos consiguen en el establecimiento de verdades matemáticas —de "juicios de verdad", como dice él. Habla como si hubiese algún género de competencia entre los seres humanos y las computadoras, y se afana en probar que ninguna computadora va a derrotarnos nunca en nuestro juego de razonamiento matemático.

Pues bien, en los primeros días de la ciencia cognitiva computacional hubo mucha gente que creyó que la tarea de esta ciencia era inventar programas que pudieran probar teoremas, descubrir hipótesis científicas, etc. Pero a estas alturas no es ese el proyecto principal de la ciencia cognitiva computacional; no pasa de ser parte del proyecto. Lo ordinario es que los científicos de esta especialidad supongan que hay un nivel de descripción del cerebro que cae entre el nivel de los procesos conscientes de pensamiento y el nivel de los procesos neurofisiológicos brutos, y procuren que sus computadoras simulen lo que acontece en la cognición en ese nivel. Por ejemplo, las teorías computacionales de la visión suelen postular que el sistema visual puede ser simulado mediante varias reglas computacionales inconscientes, a fin de dar con la forma de un objeto a partir del insumo sensorial limitado que el sistema ob-

tiene. Opino que estas teorías yerran al postular un nivel intermedio entre conciencia y neurofisiología, pero lo que importa para el presente razonamiento es que aquel proyecto sea inmune a la objeción de Penrose.

La discusión de Penrose siempre gira alrededor de los procesos conscientes de pensamiento de los matemáticos cuando están probando resultados matemáticos, y desea saber si los algoritmos demostradores de teoremas pueden dar razón de todos los éxitos de los matemáticos. No pueden, sólo que esto no es lo que se discute en la ciencia cognitiva computacional, cuando menos no en lo que concierne a la IA-débil. La cuestión es si podemos simular mediante computadoras los procesos cognocitivos humanos —sean matemáticos o no— del mismo modo como simulamos huracanes, la circulación del dinero en la economía brasileña o los procesos de combustión interna en el motor de un coche. El propósito no es obtener un algoritmo que los individuos estén tratando de seguir, sino uno que describa con exactitud lo que acontece dentro de ellos. En este proyecto simular los fracasos es tan interesante como simular los éxitos. De ahí que, por ejemplo, en los estudios computacionales sobre la visión, los modelos de cómo produce el cerebro las ilusiones de óptica sean igual de importantes que los dedicados a cómo el cerebro produce visión exacta. Aspiramos a hacernos idea de cómo funciona, no a competir con él. *En las argumentaciones de Penrose no hay nada en absoluto que milite contra un modelo computacional, así construido, del cerebro.*

Pero —dice Penrose— tales modelos computacionales no garantizan juicios de verdad. No garantizan que el sistema vaya a generar las respuestas correctas. En esto tiene razón, sólo que no viene al caso. Pues cuanta vez llegue un matemático con la respuesta correcta, habrá alguna simulación posible de sus procesos cognocitivos que modele lo que hizo, y cuanta vez llegue con la respuesta equivocada, también habrá una simulación posible. Pero las simulaciones no saben nada acerca de atinar o no, ni de alcanzar "juicios de verdad", ni nada de lo demás, ni más ni menos que las simulaciones de la economía brasileña se ocupan de cómo enriquecerse uno en Brasil.

El aspecto que es simulado del proceso tiene que ver con procesamiento inconsciente de información; el aspecto que no puede ser simulado, según la argumentación de Penrose, tiene que ver con la prueba consciente de teoremas. El hecho de que sea imposible simular los esfuerzos conscientes por probar teoremas, mediante un algoritmo demostrador, sencillamente no tiene que ver con la cuestión de si podrán ser simulados los otros aspectos de los mismísimos procesos. Las simulaciones son, nada más, esfuerzos por construir modelos de lo que realmente ocurre. De manera que la auténtica falacia de la argumentación de Penrose es

más radical que la objeción técnica que hice antes: del hecho de que no haya algoritmo demostrador de teoremas que establezca las verdades matemáticas que conocemos, sencillamente no se sigue que no haya simulación computacional de lo que acontece en nuestros cerebros cuando alcanzamos resultados matemáticos. Y la ciencia cognitiva computacional, en su versión IA-débil, no requiere nada más.

Resumiendo el punto: Penrose no logra distinguir entre algoritmos que los matemáticos estén *siguiendo* conscientemente (o, para el caso, inconscientemente), en el sentido de seguir los pasos del algoritmo, y algoritmos que *no estén siguiendo* pero que simulen o modelen con exactitud lo que sucede en los cerebros de los matemáticos cuando piensan. Nada de lo que Penrose dice muestra que sea imposible el segundo género de algoritmo. La auténtica objeción a este segundo género de algoritmo no la hace Penrose, y es que tales modelos de computadora realmente no explican nada, porque los algoritmos no desempeñan papel *causal* ninguno en el comportamiento del cerebro. No pasan de proporcionar simulaciones o modelos de lo que está ocurriendo.

La segunda objeción a Penrose es que su introducción de nociones epistémicas como "conocido" y "correcto" en la prueba original de Gödel no es tan inocente como parece creer. El problema es que la "cognoscibilidad" es parte esencial de su argumentación, sólo que no está claro qué pueda significar en este contexto.

En la segunda parte del libro, Penrose resume el actual estado de nuestro conocimiento en mecánica cuántica e intenta aplicar sus lecciones al problema de la conciencia. Buena parte de ello es difícil de leer, pero entre las exposiciones de la mecánica cuántica relativamente no técnicas que conozco, ésta me parece la más clara. Para quien alguna vez haya aspirado a enterarse acerca de la superposición, el desplome de la función de onda, la paradoja del gato de Schrödinger y los fenómenos de Einstein-Podolsky-Rosen, éste pudiera ser el mejor lugar de averiguarlo. Si bien ¿qué tiene todo ello que ver con la conciencia? La especulación de Penrose es que el mundo computable de la física clásica es incapaz de dar razón del carácter no computacional de la mente, o sea de los rasgos de la mente que a su juicio no pueden siquiera ser simulados con una computadora. Incluso una computadora mecánico-cuántica, con sus características aleatorias, no estaría en condiciones de dar razón de los aspectos esencialmente no computacionales de la conciencia humana. Pero Penrose confía en que cuando finalmente se complete la física de la mecánica cuántica, cuando dispongamos de una teoría satisfactoria de la gravedad cuántica, "nos conduciría a algo genuinamente no computable".

¿Y cómo se supone que esto funciona en el cerebro? De acuerdo con Penrose, no podemos hallar la respues-

ta al problema de la conciencia en el nivel de las neuronas porque éstas son demasiado grandes; son ya objetos explicables mediante la física clásica. Debemos escrutar la estructura interna de la neurona y encontrar allí una estructura denominada "citoesqueleto", la estructura que mantiene unida la célula y es el sistema de control para el funcionamiento de la célula. El citoesqueleto contiene diminutas estructuras como tubos, los llamados "microtúbulos", los cuales, según Penrose, desempeñan un papel decisivo en el funcionamiento de las sinapsis. He aquí la hipótesis que propone: "Según el modo de ver que provisionalmente propongo, la conciencia sería alguna manifestación de este estado citoesquelético interno, cuánticamente trabado, y de su participación en la interacción... entre niveles de actividad cuánticos y clásicos."

Dicho de otro modo, la estofa citoesquelética está enteramente revuelta con fenómenos mecánico-cuánticos, y cuando este micronivel se entrelaza con el macronivel surge la conciencia. Las neuronas no se hallan en un nivel adecuado para explicar la conciencia. Pudieran no ser sino un "dispositivo amplificador" para la acción real, residente en el nivel citoesquelético. El nivel neuronal de descripción podría ser "una simple sombra" del nivel más profundo donde tenemos que buscar el fundamento físico de la mente.

¿Qué hacer con todo esto? No objeto su carácter especulativo, ya que, tal como están las cosas, cualquier explicación de la conciencia está forzada a contener elementos especulativos. El problema con estas especulaciones es que no especulan adecuadamente acerca de cómo podríamos resolver el problema de la conciencia. Asumen la forma de: si dispusiéramos de una teoría mecánico-cuántica mejor, y si tal teoría fuese no computable, a lo mejor podríamos dar razón de la conciencia de una manera no computacional. Pero ¿cómo? No basta con decir que el misterio de la conciencia acaso se resolviera si dispusiéramos de una mecánica cuántica todavía más misteriosa que la actual. ¿Cómo se supone que funcionaría? ¿cuáles se supone que son los mecanismos causales?

Existen otros varios intentos, aparte del de Penrose, de ofrecer una explicación mecánico-cuántica de la conciencia en plan mecánico-cuántico. Suele protestarse señalando que, en efecto, semejantes planteamientos sustituyen un misterio por dos. La singularidad de Penrose es que desea agregar otro más. A los misterios de la conciencia y de la mecánica cuántica desea añadir un tercer elemento misterioso, una mecánica cuántica no computable, aún por ser descubierta.

La patente motivación de esta línea especulativa se basa en una falacia. Suponiendo que estuviese en lo cierto en que no hay nivel de descripción donde la conciencia pueda simularse, piensa que por esa razón la explicación de la conciencia debe darse en términos

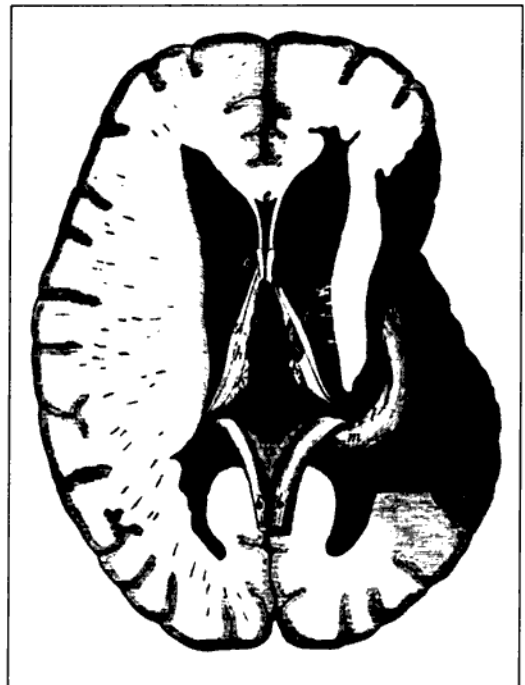
de entidades imposibles de ser simuladas. Pero lógicamente hablando esta conclusión no se sigue. De la proposición de que la conciencia no puede simularse mediante una computadora no se sigue que las entidades causantes de la conciencia tampoco puedan ser simuladas. Hay, además, una persistente confusión en la argumentación de Penrose acerca del concepto de computabilidad. Opina que, en vista de que el comportamiento de las neuronas puede ser simulado y es por lo tanto computable, cada neurona de alguna manera se torna una especie de pequeña computadora, la "computadora neural", como Penrose dice. Otra falacia. La trayectoria de una pelota de beisbol es computable y puede ser simulada con una computadora. Lo cual no convierte las pelotas en computadoras de beisbol. En resumen, hay realmente dos falacias en el meollo de su defensa de una explicación mecánico-cuántica no computacional de la conciencia. La primera es que Penrose supone que si algo no puede ser simulado mediante computadora, entonces su comportamiento no puede ser explicado por nada que pueda ser simulado; y la segunda es el supuesto tácito de que todo cuanto sea posible simular mediante computadora se torna intrínsecamente, con ello, una computadora digital.

No he explorado los presupuestos metafísicos más hondos que están detrás de toda la argumentación de Penrose. Llega armado de los recursos de la ciencia y las matemáticas, pero de hecho es un metafísico clásico, un platónico autorreconocido. Cree que vivimos en tres mundos, el físico, el mental y el matemático. Quiere mostrar cómo cada mundo fundamenta el siguiente, en círculo sin fin. Me parece que esta empresa es incoherente, si bien no intentaré mostrarlo aquí. Afirmaré nada más lo siguiente, sin argumentación: vivimos en un mundo, no dos ni tres ni veintisiete. La misión principal de una filosofía y ciencia de la conciencia, aquí y ahora, es mostrar cómo la conciencia es una parte biológica de dicho mundo, junto con la digestión, la fotosíntesis y todo lo demás. Aunque admiro enormemente tanto a Penrose como su obra, llego a la conclusión de que el principal valor de *Shadows of the Mind* es que sirve para aprender mucho acerca del teorema de Gödel y acerca de la mecánica cuántica. Acerca de la conciencia no se aprenderá gran cosa. ♪

© TNRYB



Ventrículo derecho y mitad del izquierdo del cuerpo calloso



Diseción del ventrículo lateral y el Velum Interpositum