Smolensky, P.: (próximo a aparecer) *Lectures on connectionist modeling*, Erlbaum.
Stone, G. O.: (1986) "An analysis of the delta-rule and learning statistical associations", en D. E. Rumelhart, J. L. McClelland y PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations*, Cambridge, MA., MIT Press/Bradford Books.
Sutton, R. S.: (1987) "Learning to predict by the methods of temporal differences", Technical Report 87-509, 1, GTE Laboratories, Waltham, MA.
Zadeh, L. A.: (1965) "Fuzzy sets", *Information and Control*, 8, 338-353.
Zadeh, L. A.: (1975) "Fuzzy logic and approximate reasoning", *Synthèse* 30, 407-428.
Zadeh, L. A.: (1983) "Role of fuzzy logic in the management of uncertainty in expert systems", *Fuzzy Sets and Systems*, 11, 199-227.

CAPÍTULO 15

MENTES Y CEREBROS SIN PROGRAMAS *

John Searle

El hiato

El objetivo de este capítulo es presentar una explicación provisional de algunas discusiones que he tenido con filósofos y con gente de otras disciplinas.[1] Quiero comenzar situando los puntos en cuestión en un contexto algo más amplio.

Hay una importante laguna en la vida intelectual del siglo veinte. ('Laguna' es quizás un eufemismo para 'escándalo'.) Confiamos en que podemos dar explicaciones de la conducta humana en términos ordinarios, de sentido común. Así, decimos cosas como 'Ese hombre votó por Ronald Reagan porque pensó que Reagan iba a terminar con la inflación'. Tales comentarios son parte de la psicología de sentido común o de la abuela [*common-sense or grandmother psychology*]. Para darle un nombre elegante podemos llamarla 'psicología intencionalista'['*intentionalistic psychology*']. Suponemos también que por debajo de ese nivel de explicación tiene que haber un nivel de explicación neurofisiológica. Pero, realmente, no sabemos cómo dar explicaciones neurofisiológicas

* "Minds and Brains Without Programs", en *Mindwaves*, Colin Blakemore y Susan Greenfield (comps.), Oxford, Blackwell, 1989, págs. 209-233. Con autorización del autor y de *Basil Blackwell*.

1. Este capítulo está basado en una conferencia dada en Oxford cuando estaba en Inglaterra para grabar las conferencias Reith de 1984 para la BBC, conferencia que fue hecha originalmente para una publicación separada. Hay una superposición considerable entre el material de las conferencias Reith y el material de esta conferencia. Las conferencias han sido publicadas como Minds, Brains and Science (BBC Publications, 1984; Harvard Univ. Press, 1984). Pido disculpas a los que escucharon y leyeron las conferencias Reith por la repetición. Publico este artículo en forma separada, en parte porque Colin Blakemore y Susan Greenfield me convencieron de que sería una contribución útil para este volumen, a pesar de repetir material publicado en otra parte, y en parte porque me da la oportunidad de expandir y explicar varios puntos que se expusieron en las conferencias Reith.

de la conducta humana ordinaria. No sabemos cómo hacer aseveraciones tales como 'El hombre votó por Reagan debido a cierta condición en su tálamo'. Esto nos coloca en una situación intelectual embarazosa. Usamos con razonable confianza la psicología de la abuela en el nivel más elevado y pensamos que tiene que haber una ciencia dura sustentándola en el nivel más bajo, pero no tenemos la más vaga idea de cómo funciona el nivel para explicar los casos específicos de conducta humana normal. Usamos la psicología de la abuela todo el tiempo, pero nos avergüenza llamarla ciencia. Nadie, por ejemplo, tiene la presencia de ánimo como para presentarse a la National Science Foundation a pedir una beca para hacer psicología de la abuela. Sin embargo, no sabemos lo suficiente acerca del nivel más bajo como para hacerlo funcionar. Parece, pues, que tenemos un hiato.

$$\frac{\text{psicología intencionalista}}{\text{neurofisiología}} \quad \leftarrow \text{hiato}$$

Algunos de los grandes esfuerzos intelectuales del siglo veinte han sido intentos de salvar el hiato, de encontrar algo que fuera una ciencia de la conducta humana pero que no fuera psicología de sentido común ni tampoco fuera neurofisiología. Y si uno vive lo suficiente, es interesante mirar atrás y ver los cadáveres de teorías que supuestamente iban a salvar el hiato. Durante mi vida el fracaso más espectacular fue el del conductismo. Pero también viví otros esfuerzos fallidos. Hubo la teoría de juegos y la teoría de la información. No creo que nadie que lea esto sea tan viejo como para no recordar la cibernética, pero en un tiempo se hicieron grandes aseveraciones acerca del futuro de la cibernética. Hubo algo llamado 'estructuralismo', que fue seguido por algo llamado 'post-estructuralismo'. Ahora está la sociobiología, otro candidato para salvar el hiato.

Sin embargo, el principal candidato se llama hoy 'ciencia cognitiva', y con frecuencia se piensa que el programa central de investigación en la ciencia cognitiva es la inteligencia artificial. Hay diferentes escuelas de ciencia cognitiva y de inteligencia artificial, pero la teoría más ambiciosa para salvar el hiato es la que dice que la investigación en psicología cognitiva y en inteligencia artificial ha establecido que la mente es al cerebro como el programa del computador es al *hardware* del computador. La siguiente ecuación es muy común en la literatura: mente/cerebro = programa/*hardware*. Para distinguir este punto de vista de versiones más cautelosas de la inteligencia artificial, lo he llamado 'inteligencia artificial fuerte' ('IA fuerte', para abreviar). De acuerdo con la IA fuerte, un computador adecuadamente programado, con los *inputs* y *outputs* correctos, tiene literalmente una mente en el mismo sentido en que usted y yo la tenemos.

Este punto de vista tiene algunas consecuencias interesantes. Tiene la consecuencia, por ejemplo, de que no hay nada esencialmente biológico respecto de la mente humana. Sucede que los programas que son constitutivos de las mentes son operados [*are run*] en el *wetware* que tenemos en nuestra máquina biológica, en el computador biológico de la cabeza. Pero esos mismos programas podrían ser operados en el *hardware* de cualquier computador que fuera capaz de sostener el programa. Y esto tiene la consecuencia adicional de que cualquier cosa, cualquier sistema, podría tener pensamientos y sentimientos —y no sólo *podría tener*, sino que *tendría que tener* pensamientos y sentimientos— exactamente en el mismo sentido en que nosotros los tenemos, con la sola condición de que se esté operando el programa correcto. Esto es, si se tiene el programa correcto, con los *inputs* y *outputs* correctos, entonces cualquier sistema que opere ese programa, al margen de su estructura química (sea que esté hecho de latas de cerveza viejas o de chips de siliconas o de cualquiera otra substancia) *tiene que tener* pensamientos y sentimientos, exactamente de la misma manera en que usted y yo los tenemos. Y esto es así porque en eso consiste tener una mente: en tener el programa correcto. Ahora bien, siempre que ataco este punto de vista, muchos dicen: 'Pero seguramente nadie puede creer eso'. Voy a decirles los nombres de algunas personas que creen eso, así no piensan que estoy atacando a un personaje imaginario.

Herbert Simon, de la Carnegie-Mellon University, ha escrito en numerosas ocasiones que ya contamos con máquinas que pueden pensar en un sentido literal; que pueden pensar en el mismo sentido que usted y yo lo hacemos. Los filósofos han discutido durante siglos si se puede o no construir una máquina que piense, y ahora lo hacen a diario en Carnegie-Mellon. Allan Newell, el colega de Simon, en una conferencia que le escuché dar en San Diego en la reunión fundacional de la Cognitive Science Society, dijo que se había descubierto (no que era alguna hipótesis que se estaba considerando, sino que se había 'descubierto') que la inteligencia es exclusivamente manipulación de símbolos físicos. De modo que cualquier máquina que sea capaz de manipular los símbolos correctos de manera correcta, tiene procesos inteligentes, exactamente en el mismo sentido en que usted y yo los tenemos. Marvin

Minsky dice que la próxima generación de computadores va a ser tan inteligente que vamos a tener suerte si nos dejan en casa como mascotas. Y Freeman Dyson es citado en el *New York Times* como habiendo dicho que dado que ahora sabemos que procesos mentales tales como la conciencia son procesos puramente formales, hay una ventaja evolutiva en tener tales procesos formales (conciencia y demás) funcionando en chips de siliconas y alambres, porque en un universo que se está enfriando, ese tipo de material tiene más capacidad para sobrevivir que los organismos como los nuestros, hechos de una desaliñada maquinaria biológica. Así que el próximo paso evolutivo, según este punto de vista, va a estar hecho de alambres y siliconas. En esta literatura mi texto favorito (y les recomiendo esta literatura porque es maravillosa) es de John McCarthy, el inventor del término 'inteligencia artificial'. McCarthy escribió: "Puede decirse que máquinas tan simples como los termostatos tienen creencias...". Y agregó, por cierto: "Tener creencias parece ser una característica de la mayoría de las máquinas capaces de resolver problemas".[2] De modo que le pregunté: "John, ¿qué creencias tiene tu termostato?". Admiro su coraje. Dijo: "Mi termostato tiene tres creencias. Mi termostato cree que hace demasiado calor aquí, que hace demasiado frío aquí y que la temperatura es adecuada aquí".

Bien, me gusta esta tesis por una simple razón. La ecuación mente/cerebro = programa/*hardware*, no es usual en filosofía porque es una tesis razonablemente clara. Uno la puede enunciar con razonable precisión. Y a diferencia de la mayoría de las tesis filosóficas, está sujeta a una muy simple y, pienso, decisiva refutación. He publicado la refutación en otra parte, pero la repetiré brevemente aquí porque en la comunidad de la inteligencia artificial no se acepta universalmente que haya refutado, de hecho, ese punto de vista. Después quiero tocar una cuestión más profunda, que es ésta: una de las razones por las que la gente cree en la IA fuerte es que no puede ver otra manera de resolver el problema mente-cuerpo. Estoy convencido de que una de las fuentes de la creencia de que tener una mente equivale a tener un programa de computación, es que la gente no puede ver otra forma de resolver el problema mente-cuerpo sin recurrir al dualismo. Con frecuencia se me pregunta: "Bien, si no acepta el análisis de la mente que ofrece la inteligencia artificial, entonces, ¿cuál es su solución al problema mente-cuerpo? ¿No está usted forzado a caer en el dualismo o en el misticismo o en el vitalismo o en algún enfoque, igualmente misterioso?". Así que realmente tengo dos tareas. Quiero refutar a la inteligencia artificial fuerte y quiero resolver el problema mente-cuerpo.

2. McCarthy, John (1979): "Ascribing mental qualities to machines", Stanford Artificial Intelligence Laboratory Memo AIM-326, pág. 2, *Computer Science Department Report*, No. STAN-CS-79-725, marzo de 1979.

Una nueva visita a la habitación china

El argumento en contra de la IA fuerte es, me temo, bastante simple. El argumento se me ocurrió cuando leí el libro de Schank y Abelson acerca de sus programas de comprensión-de-relatos [*story-understanding programs*].[3] Algunos estarán familiarizados con ellos; pero volveré a repasar cómo funcionan sus programas. Se trata de programas muy ingeniosos que han sido diseñados en Yale University. Los programas realizan lo que ellos llaman 'comprender relatos'. El computador recibe un relato muy simple como *input*. Una historia típica sería la que sigue:

'Un hombre fue a un restaurante y pidió una hamburguesa. Cuando le trajeron la hamburguesa estaba muy quemada. El hombre salió enfurecido del restaurante sin haber pagado la hamburguesa.'

Entonces uno le pregunta al computador: '¿Comió el hombre la hamburguesa?'. Y el computador dice como *output*: 'No, el hombre no comió la hamburguesa'. O uno le da a la computadora otro relato:

'Un hombre fue a un restaurante y pidió una hamburguesa. Cuando le trajeron la hamburguesa se deleitó con ella, y cuando dejó el restaurante pagó la cuenta y dejó una buena propina al mozo'.

Si entonces uno le pregunta al computador: '¿Comió el hombre la hamburguesa?'. El computador responde: 'Sí, el hombre comió la hamburguesa'. Nótese que ninguna de las dos historias decía explícitamente si el hombre había comido o no la hamburguesa. ¿Cómo funciona? Funciona porque el programa tiene en su base de datos lo que se llama un 'guión-de-restaurante' [*restaurant script*]. El guión-de-restaurante es la representación de cómo son las cosas normalmente en los restaurantes. Cuando el computador toma el relato, lo aparea con el guión-de-restaurante, y entonces, cuando toma la pregunta acerca del relato, aparea la pregunta con ambos: el relato y el guión-de-restaurante. Dado que 'sabe' cómo se supone que suceden las cosas en los restaurantes, puede producir la respuesta adecuada. La alegación, que se hace con frecuen-

3. Schank, R.C. y Abelson, R.P. (1977): *Scripts, Plans, Goals and Understanding*, Hillsdale, NJ., Erlbaum.

cia respecto de los programas es que como la máquina satisface el test de Turing, la máquina tiene que comprender literalmente el relato.[4] Tiene que comprender literalmente el relato en el mismo exacto sentido en que usted y yo comprenderíamos esos relatos si nos hicieran tales preguntas y diéramos buenas respuestas.

Me parece que hay una refutación muy simple a esa alegación. La refutación consiste en imaginar que uno es la máquina. A mí me gusta imaginarla de la siguiente manera.

Supóngase que estoy encerrado en una habitación. En esa habitación hay dos grandes cestos llenos de símbolos chinos, junto con un libro de reglas en español acerca de cómo aparear los símbolos chinos de una de las cestas con los símbolos chinos de la otra cesta. Las reglas dicen cosas como: 'Busque en la canasta 1 y saque un signo garabateado, y póngalo al lado de un cierto signo garabateado que saque de la canasta 2'. Adelantándonos un poco, esto se llama una 'regla computacional definida en base a elementos puramente formales'. Ahora bien, supóngase que la gente que está fuera de la habitación envía más símbolos chinos junto con más reglas para manipular y aparear los símbolos. Pero esta vez sólo me dan reglas para que les devuelva los símbolos chinos. Así que estoy aquí, en mi habitación china, manipulando estos símbolos. Entran símbolos y yo devuelvo los símbolos de acuerdo con el libro de reglas. Ahora bien, sin yo saberlo, quienes organizan todo esto fuera de la habitación llaman a la primera cesta un 'guión-de-restaurante' y a la segunda cesta un 'relato acerca del restaurante', a la tercera hornada de símbolos la llaman 'preguntas acerca del relato,' y a los símbolos que les devuelvo 'respuestas a las preguntas'. Al libro de reglas lo llaman 'el programa', ellos se llaman 'los programadores' y a mí me llaman 'el computador'. Supóngase que después de un tiempo soy tan bueno para responder esas preguntas en chino que mis respuestas son indistinguibles de las de los nativos hablantes del chino. Con todo, hay un punto muy importante que necesita ser enfatizado. Yo no comprendo una palabra del chino, y no hay forma de que pueda llegar a entender el chino a partir de la instanciación de un programa de computación, en la manera en que la describí. Y éste es el quid del relato: *si yo no comprendo chino en esa situación, entonces tampoco lo comprende ningún otro computador digital sólo en virtud de haber sido adecuadamente programado, porque ningún computador digital por el solo hecho de ser un computador digital, tiene algo que yo no tenga.* Todo lo que tiene un

4. No quiero implicar que Schank o Abelson hagan esta afirmación.

computador digital, por definición, es la instanciación de un programa formal de computación. Pero como yo estoy instanciando el programa, como suponemos que tenemos el programa correcto con los *inputs* y *outputs* correctos, y yo no comprendo el chino, entonces no hay forma de que cualquier otro computador digital *sólo en virtud de instanciar el programa* pueda comprender el chino.

Éste es el núcleo central del argumento. Pero su quid, pienso, se perdió en el montón de bibliografía desarrollada subsiguientemente a su alrededor; así que quiero enfatizarlo. El quid del argumento no es que de una u otra manera tenemos la 'intuición' de que no comprendo el chino, de que me *inclino a decir* que no comprendo el chino pero que, quien sabe, quizá realmente lo entienda. Ese no es el punto. El quid del relato es recordarnos una verdad conceptual que ya conocíamos, a saber, que hay diferencia entre manipular los elementos sintácticos de los lenguajes y realmente comprender el lenguaje en un nivel semántico. Lo que se pierde en la *simulación del* comportamiento cognitivo de la IA, es la distinción entre la sintaxis y la semántica.

El quid del relato puede enunciarse ahora más genéricamente. Un programa de computación, por definición, tiene que ser definido de manera puramente sintáctica. Es definido en términos de ciertas operaciones formales realizadas por la máquina.[5] Eso es lo que hace del computador digital un instrumento tan poderoso. Uno y el mismo sistema de *hardware* puede instanciar un numero indefinido de programas de computación diferentes, y uno y el mismo programa de computación puede operarse en *hardwares* diferentes, porque el programa tiene que ser definido de manera puramente formal. Pero por esa razón la simulación formal de la comprensión del lenguaje nunca va a ser en sí lo mismo que la duplicación. ¿Por qué? Porque en el caso de comprender realmente un lenguaje, tenemos algo más que un nivel formal o sintáctico. Tenemos la semántica. No manipulamos meramente símbolos formales no interpretados, sabemos realmente qué significan.

Esto puede mostrarse enriqueciendo un poco el argumento. Estoy allí, en la habitación china, manipulando esos símbolos chinos. Supóngase ahora que algunas veces los programadores me dan relatos en español y que me hacen preguntas, también en castellano, acerca de esos

5. Los programas de 'Comprensión lingüística' ['*Language understanding*'] tienen característicamente una 'sintaxis', una 'semántica' y en algunos casos hasta una 'pragmática'. Por supuesto, esto es irrelevante para el argumento, porque los tres niveles son computacionales, esto es, 'sintácticos', en el sentido en que estoy usando ahora la palabra.

relatos. ¿Cuál es la diferencia entre los dos casos? Tanto en el caso del español, como en el caso del chino, satisfago el test de Turing. Esto es, doy respuestas que son indistinguibles de las respuestas que daría un hablante nativo. En el caso del chino lo hago porque los programadores son buenos para diseñar el programa, y en el caso del español porque soy un hablante nativo. ¿Cuál es la diferencia, entonces, si mi actuación es equivalente en los dos casos? Me parece que la respuesta es obvia. La diferencia es que sé español. Sé qué significan las palabras. En el caso del español no sólo tengo una sintaxis, tengo una semántica. Atribuyo un contenido semántico o significado a cada una de esas palabras, y por lo tanto estoy haciendo más que lo que un computador digital puede hacer en virtud de instanciar un programa. Tengo una interpretación de las palabras y no sólo de los símbolos formales. Nótese que si tratamos de dar al computador una interpretación de los símbolos formales lo único que podemos hacer es darle más símbolos formales. Todo lo que podemos hacer es poner más símbolos formales no interpretados. Por definición, el programa es sintáctico, y la sintaxis por sí misma nunca es suficiente para la semántica.[6]

Bueno, en esto consiste mi rechazo de la ecuación, mente/cerebro = programa/*hardware*. Instanciar el programa correcto nunca es suficiente para tener una mente. Tener una mente es algo más que instanciar un programa de computación. Y la razón es obvia. Las mentes tienen contenidos mentales. Tienen contenidos semánticos Así como tienen un nivel sintáctico de descripción.

Hay un mal entendido persistente acerca de mi argumento, que quiero neutralizar ya. Algunos suponen que sostengo que en principio es imposible para los chips de siliconas duplicar el poder causal del cerebro. Ése no es mi argumento; es más, no tiene ninguna conexión con mi argumento. Que los poderes causales de las neuronas puedan ser duplicados en algún otro material, como chips de siliconas, válvulas electrónicas, transistores, latas de cerveza, o alguna desconocida substancia química, es una cuestión fáctica, que no ha de ser resuelta apelando a bases puramente filosóficas o a priori. El quid de mi argumento es que uno no puede duplicar los poderes causales del cerebro sólo en virtud

6. Alguna gente verdaderamente temeraria de la IA ha propuesto que no soy yo quien comprende sino la habitación entera, esto es, el sistema que me contiene, el programa, los canastos, la ventana al exterior, etcétera. Pero esta respuesta está sujeta a la misma objeción. Así como yo no tengo manera de pasar de la sintaxis a la semántica, tampoco lo tiene el sistema total. El sistema total no tiene manera de saber qué significa cualquiera de los símbolos formales.

de instanciar un programa de computación, debido a que el programa de computación tiene que ser definido de manera puramente formal. Es importante enfatizar que la inteligencia artificial, sea fuerte o de otra manera, no tiene nada que ver con las propiedades químicas de la silicona o de cualquiera otra sustancia. Una vez que el partidario de la IA concede que tales propiedades son relevantes, ha abandonado la tesis de la IA. La IA es acerca de los poderes 'cognitivos' de los programas. Nada tiene que ver con las propiedades químicas específicas de las realizaciones de los programas en el *hardware*.

Sin embargo, esto nos enfrenta a la segunda cuestión. Si rechazamos la ecuación y rechazamos la IA como salvando el hiato, entonces, ¿cuál es nuestro análisis de la relación entre el nivel de la intencionalidad y el nivel de la neurofisiología? Una respuesta breve es que la razón de que lo que cierra el hiato siempre falle, es que no hay ningún hiato que salvar. No hay hiato alguno entre el nivel de las explicaciones intencionalistas y el nivel de las explicaciones neurofisiológicas. Pero para fundamentar esto necesito, como prometí antes, resolver el problema mente-cuerpo.

Cuatro enigmas

Antes de considerar directamente el problema mente-cuerpo, quiero volver atrás por un momento y preguntar por qué este problema parece ser tan dificultoso. ¿Por qué en filosofía, psicología y neurofisiología, todavía tenemos el problema mente-cuerpo? Desde Descartes, al menos, la forma general del problema mente-cuerpo ha sido el problema de reconciliar nuestras creencias de sentido común y precientíficas acerca de la mente con nuestra concepción científica de la realidad. Nuestra concepción científica del mundo, como un sistema físico o como un conjunto de sistemas físicos en interacción, ha crecido en poder y comprehensividad, y parece cada vez más difícil encontrar en esa concepción un lugar para la mente. Algunos de los puntos de vista precientíficos que parecen ser cuestionados por el crecimiento de una visión científica del mundo derivan de la religión o de la moral —doctrinas tales como la inmortalidad del alma, el libre albedrío, la naturaleza de la responsabilidad moral—, y acerca de esas cuestiones no tendré nada que decir en esta discusión. Me detendré en una pregunta más restringida y, creo, más apremiante: ¿cómo podemos encuadrar lo que sabemos o lo que creemos saber acerca del mundo en general, con lo que

sabemos o lo que creemos saber acerca del funcionamiento de nuestras propias mentes? Dejando a un lado las especulaciones de la religión y las presuposiciones de la moral, sabemos una cantidad de cosas acerca de nuestras mentes, y mi objetivo es dar una explicación coherente de las relaciones entre lo que sabemos acerca de nuestras propias mentes y lo que sabemos acerca de la forma en que funciona el mundo en general. ¿Por qué este problema más restringido, no religioso y no moral, ha sido tan dificultoso? ¿Por qué, para decirlo una vez más, hay todavía un problema mente-cerebro o mente-cuerpo?

Las características [*features*] de nuestra concepción de sentido común de la mente que parecen difíciles de asimilar a nuestra concepción científica general del mundo son, al menos, las cuatro siguientes:

Conciencia

Yo, en el momento de escribir esto, y usted, en el momento de comprenderlo, somos conscientes. Que el mundo contiene estados mentales conscientes es un hecho evidente, pero es difícil ver cómo meros sistemas físicos pueden tener conciencia. ¿Cómo puede ocurrir tal cosa? ¿Cómo, por ejemplo, puede ser consciente este trozo de materia gris y blanca que está adentro de mi cráneo?

Intencionalidad

Muchos de mis estados mentales, como por ejemplo, mis creencias y deseos y mis percepciones visuales e intenciones, están dirigidas a, o son acerca de objetos y estados en el mundo, distintos de ellos mismos. Este rasgo, llamado 'intencionalidad' es una característica de las mentes humanas. Pero, nuevamente, ¿cómo puede ocurrir una cosa tal? ¿Cómo pueden ser *acerca* de algo procesos en mi cerebro que, después de todo, consisten finalmente en 'átomos en el vacío' ['*atoms in the void*']? ¿Cómo pueden átomos en el vacío *representar* algo? Uno se inclina a decir: las cosas y los procesos en el mundo simplemente son; sea que pensemos en *procesos* como la digestión y en secuencias de neuronas excitadas o en *cosas* físicas corrientes como piedras y árboles, parece imposible que alguna de ellas pueda ser *acerca de* algo. ¿Cómo puede el *acerca de algo* [*aboutness*] ser un rasgo intrínseco del mundo?

Subjetividad

Los estados mentales son característicamente subjetivos. Pero es difícil entender cómo el mundo físico objetivo, igualmente abierto a todos los observadores competentes, puede contener algo esencialmente subjetivo como, por ejemplo, estados mentales conscientes. Interpretada ingenuamente, la subjetividad de los estados mentales está marcada por hechos tales como que yo tengo mis estados y no los suyos, que mis estados me son accesibles de una manera en que no son accesibles para usted; percibo el mundo desde mi punto de vista y no desde su punto de vista, etcétera. ¿Cómo puede la subjetividad ser una parte real del mundo?

Causación intencional

Aun si hubiera cosas tales como estados mentales, es difícil ver cómo ellos podrían producir una diferencia real en el mundo. ¿Podría algo, por decirlo de alguna manera, tan 'gaseoso' y 'etéreo' como un estado mental consciente tener algún impacto en un objeto físico como el cuerpo humano? ¿Cómo podrían los fenómenos mentales empujar objetos o tener cualquiera otra significación física? ¿No serían los eventos mentales, si existieran, sólo epifenoménicos?

Llamemos a estos problemas, respectivamente, los problemas de la conciencia, de la intencionalidad, de la subjetividad y de la causación intencional. Aunque no todos los estados mentales tienen estas cuatro características, son sin embargo características reales y típicas de los fenómenos mentales. Sabemos, por ejemplo, que las personas están a menudo en un estado de conciencia, que tienen con frecuencia, por ejemplo, pensamientos y sentimientos que refieren a objetos y estados de cosas fuera de ellas mismas, que aprehenden el mundo desde un punto de vista subjetivo, y que sus pensamientos y sentimientos tienen relevancia en su comportamiento. Creo que cualquier explicación acerca del problema mente-cerebro debe ser capaz, al menos, de dar cuenta de todos estos hechos.

En el punto de vista que se adopta en este ensayo acerca de los estados mentales, ellos y los procesos mentales son fenómenos biológicos reales en el mundo, tan reales como la digestión, la fotosíntesis, la lactancia o la secreción de bilis. El objetivo de este capítulo no es mostrar en detalle cómo tales fenómenos biológicos están relacionados con los procesos neurofisiológicos del cerebro —nadie sabe en detalle cómo

están relacionados—, su objetivo, en cambio, es más modesto: mostrar cómo es posible que los estados mentales puedan ser fenómenos biológicos en el cerebro. Creo que un supuesto típico pero no enunciado de muchas de las implausibles doctrinas contemporáneas acerca de la mente —doctrinas tales como el conductismo o la inteligencia artificial fuerte— es que es sencillamente imposible acomodar una explicación ingenua de sentido común acerca de la mente con una visión científica del mundo. Y creo que es la desesperación causada por el sentimiento de que no puede darse ninguna explicación coherente que acomode el mentalismo de sentido común con la ciencia dura, lo que ha llevado a la gente a decir las cosas implausibles y algunas veces tontas que se dicen acerca de la naturaleza de la mente. El punto de vista que voy a exponer acerca de la relación de la mente y el cerebro, es consistente con lo que se sabe acerca del funcionamiento del cerebro y es también consistente con un enfoque biológico general de los fenómenos biológicos. Mi enfoque no intenta tratar a la mente como algo formal o abstracto, tal como hace la IA fuerte, ni tampoco intenta tratar a la mente simplemente como un conjunto neutral de poderes causales sin características mentales intrínsecas, tal como hacen ciertas formas de funcionalismo. Francamente pienso que el enfoque que voy a presentar es más bien un punto de vista obvio y de sentido común, y hasta que me vi envuelto en esas polémicas recientes, suponía que era ampliamente aceptado, tan ampliamente aceptado que hasta no merecía una enunciación expresa. Sin embargo, mis formulaciones previas han sido calificadas por mis críticos de 'místicas' (Ringle),[7] 'sofísticas' (Dennett),[8] 'religiosas' (Hofstadter),[9] etcétera. Quizás, entonces, valga la pena explicar la posición con algún detalle para que se pueda ver que esos cargos son realmente infundados. No preciso enfatizar que no soy la primera persona en sostener ese punto de vista y que similares enfoques biológicos del problema mente-cuerpo pueden encontrarse al menos tan atrás como el siglo diecinueve.

7. Ringle, Martin (1980): "Mysticism as a philosophy of artificial intelligence", comentario sobre "Minds, brains and programs" de Searle, *The Behavioral and Brain Sciences*, 3: 444.

8. Dennett, Daniel (1980): "The milk of human intentionality", comentario sobre "Minds, brains and programs" de Searle, *The Behavioral and Brain Sciences*, 3: 428.

9. Hofstadter, Douglas R. (1980): "Reductionism and religion", comentario sobre "Minds, brains and programs", *The Behavioral and Brain Sciences*, 3: 433.

El cerebro y su mente

¿Cómo funciona el cerebro? En detalle, nadie lo sabe. Yo tengo una ignorancia de *amateur* respecto de este tema, pero aun los mejores expertos están confundidos, hasta ahora, respecto de lo que uno pensaría que son las preguntas más fundamentales. ¿Cuál es exactamente la neurofisiología de la conciencia? ¿Por qué necesitamos dormir? ¿Cómo se almacenan los recuerdos en el cerebro, con exactitud? ¿Por qué el alcohol nos emborracha? ¿Por qué la aspirina alivia el dolor? Recientemente, en 1978, un neurólogo famoso, David Hubel, escribió: 'Hay [áreas del cerebro] del tamaño de un puño, de las cuales se puede decir que estamos casi en el mismo estado de conocimiento del que estábamos con relación al corazón antes de darnos cuenta de que bombeaba sangre'.[10] Más aún, en nuestra ignorancia, buscamos a tientas metáforas y analogías, generalmente basadas en la última tecnología. Así, hoy en día, el punto de vista de moda es que el cerebro es un computador digital, pero en mi niñez se aseguraba que era un tipo de tablero de distribución telefónico; Charles Sherrington comparó al cerebro con un sistema de telégrafo y con un telar de *jacquard*; Sigmund Freud lo comparó con bombas hidráulicas y sistemas electromagnéticos; Leibniz con un molino, y me han dicho que ciertos griegos antiguos pensaban que el cerebro funcionaba como una catapulta. El último punto de vista entre los neurofisiólogos es que el cerebro funciona como un sistema de selección natural darwiniano.

Sin embargo, aunque haya mucho para aprender, no somos totalmente ignorantes, y en una discusión como ésta necesitamos recordar unas pocas cosas elementales acerca del cerebro. Como todos los órganos, el cerebro consiste en células. Sin embargo, a diferencia de los demás órganos, el cerebro y el resto del sistema nervioso consiste, en gran parte, en un tipo muy especial de células: las neuronas. Las estimaciones corrientes dicen que hay entre 50 y 100 mil millones de neuronas en el cerebro humano. Hay una gran diversidad de tipos de neuronas, pero la neurona típica consiste en un cuerpo celular o soma, con dos tipos de fibras largas que emergen de él, un único axón y una cantidad de dendritas. Las neuronas se ponen en contacto unas con otras en ciertas protuberancias pequeñas llamadas sinapsis. Los axones y las dendritas realmente no se funden en las sinapsis; el axón tiene, como caracte-

10. Hubel, D. (1978): "Vision and the brain", *Bulletin of the American Academy of Arts and Sciences*, abril de 1978, 31, No. 7, 18.

rística, una pequeña protuberancia, el botón [*bouton*], que toca en la punta la dendrita, y la pequeña brecha que queda entre ellos es el espacio sináptico [*synaptic cleft*]. También hay sinapsis en el soma. Algunas neuronas en el cerebelo tienen hasta 200.000 sinapsis en una célula. Una de las funciones básicas de la neurona es la transmisión de impulsos eléctricos, esto es, cambios 'todo-o-nada' en el potencial eléctrico. Cada impulso eléctrico pasa del soma a lo largo del axón. Sin embargo, en la mayoría de las neuronas el impulso eléctrico no pasa directamente de una neurona a la siguiente; más bien, el impulso eléctrico, cuando llega al botón, causa la liberación de pequeñas cantidades de fluido de los pequeños compartimientos del botón, las vesículas sinápticas, al espacio sináptico. La liberación de esos fluidos (los neurotransmisores) en las sinapsis, puede tener un efecto excitatorio o inhibitorio en la neurona siguiente. Si es excitatorio, va a tender a causar el disparo [*firing*] de la neurona siguiente o va a incrementar el índice de disparo [*rate of firing*]. Si es inhibitorio, va a tender a impedir que la neurona se dispare o a disminuir el índice de disparo. Desde un punto de vista funcional lo importante no es que la neurona dispare, porque de todas maneras muchas neuronas disparan permanentemente. Lo que es importante son las variaciones del *índice* de disparo de las neuronas; específicamente, las variaciones en el índice de disparos de los axones respecto de la suma de todas las excitaciones e inhibiciones en las dendritas.

Es importante enfatizar este punto porque muchos autores han supuesto, erróneamente, que el carácter 'todo-o-nada' del disparo de los impulsos nerviosos constituye una prueba de que los principios del funcionamiento del cerebro son los de un computador digital.[11] Nada puede estar más alejado de la verdad. Hasta donde sabemos, el aspecto funcional de la neurona es la variación no-digital en el índice de disparo.

En la descripción tradicional del cerebro, es decir, la descripción que toma a la neurona como la unidad fundamental de funcionamiento del cerebro, lo más importante acerca de la relación entre el cerebro y la mente es simplemente esto. Toda la enorme variedad de *inputs* que recibe el cerebro —los fotones que alcanzan la retina, las ondas sonoras que estimulan las células del oído interno, la presión en la piel que activa las terminales nerviosas correspondientes a presión, calor, frío y dolor, etcétera— todos estos *inputs* son convertidos a un medio común: índices variables de disparos de neuronas. Más aún, y de manera igualmente notable, esos índices variables de disparos de las neuronas relativos a diferentes circuitos neuronales y a diferentes condiciones locales en el cerebro, producen toda la variedad y heterogeneidad de la vida mental del agente humano o animal. El aroma de una rosa, la experiencia del azul del cielo, el gusto de las cebollas, el pensamiento de una fórmula matemática, todo esto es producido por índices variables de disparos de neuronas, en circuitos diferentes relativos a condiciones locales diferentes del cerebro. Ahora bien, ¿qué son exactamente esos circuitos neuronales diferentes y cuáles son los entornos locales diferentes que dan cuenta de las diferencias en nuestra vida mental? En detalle nadie lo sabe, pero tenemos buena prueba de que ciertas regiones del cerebro se especializan en cierto tipo de experiencias. El córtex visual juega un rol especial en las experiencias visuales, el córtex auditivo en experiencias auditivas, etcétera. La visión es una de las funciones del cerebro mejor comprendida (o menos inadecuadamente comprendida), y en el caso de la visión parece haber neuronas muy especializadas en el córtex visual capaces de responder a diferentes rasgos específicos de los estímulos visuales. Supóngase que estímulos auditivos alimentaran al córtex visual y que estímulos visuales alimentaran al córtex auditivo. ¿Qué sucedería? Hasta donde sé, nadie ha realizado jamás ese experimento, pero parece razonable suponer que el estímulo auditivo sería 'visto', esto es, que produciría experiencias visuales, y que los estímulos visuales serían 'escuchados', esto es, producirían experiencias auditivas, debido en ambos casos a rasgos específicos aunque ampliamente desconocidos del córtex visual y del auditivo, respectivamente. Aunque esta hipótesis es especulativa, tiene algún soporte independiente si se piensa que un golpe en el ojo produce un *flash* visual ('ver las estrellas'), aun cuando no sea un estímulo óptico.

En mi visión de lego la cantidad de conocimiento que tenemos actualmente acerca de la naturaleza y el funcionamiento de las neuronas, es bastante impresionante: sin embargo, existe ahora fuerte evidencia de que para entender el papel del cerebro en la vida mental necesitamos entender el funcionamiento del cerebro en niveles más elevados que el de las neuronas individuales, y, en particular, que en esos niveles elevados necesitamos entender el funcionamiento de los sistemas de neuronas organizadas en redes neurales o en circuitos neurales. Para muchas funciones del cerebro la unidad de funcionamiento no es la célula singular sino la red de células y, en ese nivel, el funcionamiento del

11. Oppenheim, Paul y Putnam, Hilary (1958): "Unity of science as a working hypothesis", en Feigl, Scriven y Maxwell (comps.), *Minnesota Studies in the Philosophy of Science*, vol. 2, *Concepts, Theories, and the Mind-Body Problem*, pág. 19, Minneapolis, Univ. of Minnesota Press.

cerebro consiste en la interacción entre un gran conjunto de redes neurales. La mejor prueba anatómica en favor de la existencia de redes que funcionan más o menos independientemente, es la existencia de módulos neurales en la forma de columnas verticalmente orientadas, de cilindros o de lajas [*slabs*] de células en el córtex.[12] Como dice Gerald Edelman, 'El mayor logro en el pensamiento acerca del córtex, la mayor revolución, es que el córtex no es una hoja [*sheet*] continua dispuesta horizontalmente, sino que está verticalmente organizado como pilas de lajas o columnas'.[13] Estos módulos pueden variar en la cantidad de neuronas que contienen, desde 50 hasta 50.000 o aún más. Desde el punto de vista modular, la significación actual de la neurona reside en la contribución que hace al funcionamiento del módulo.

No sé si la combinación de la explicación neuronal y de la explicación modular o quizá de algún otro tipo de explicación, es la explicación correcta del funcionamiento del cerebro. Pero una conclusión surge con claridad aun de la más corriente investigación del funcionamiento del cerebro: *los fenómenos mentales, concientes o inconcientes, visuales o auditivos, los dolores, cosquilleos, picazones, pensamientos, y el resto de nuestra vida mental, son causados por procesos que suceden en el cerebro*. Los fenómenos mentales son un resultado de los procesos electroquímicos en el cerebro, tanto como la digestión es el resultado de procesos químicos que suceden en el estómago y en el resto del aparato digestivo. Creo que éste es un hecho obvio acerca de cómo funciona el mundo y sin embargo sus implicaciones completas generalmente no son percibidas por los estudiosos de la inteligencia artificial, la ciencia cognitiva o la filosofía. También es importante enfatizar que los procesos causales relevantes son enteramente internos al cerebro. Aunque *de hecho* los eventos mentales median entre los estímulos externos y las respuestas motoras, no hay una conexión *esencial*. Un hombre puede, por ejemplo, tener un dolor terrible sin tener un estímulo de dolor [*pain stimulus*] en los nervios periféricos o un comportamiento correspondiente a dolor [*pain behavior*]. Este simple hecho es suficiente para desacreditar toda la tradición conductista en filosofía.

Para substanciar esto, un poco al menos, contemos una parte del

12. Véase J. Szentagothai, "The Brain-Mind Relation: A Pseudoproblem?", en la compilación a la que pertenece este trabajo.

13. Edelman, Gerald (1982): "Through a computer darkly: group selection and higher brain function", *Bulletin of the American Academy of Arts and Sciences*, octubre de 1982, 36, Nó. 1, 28.

relato causal que corresponde a un tipo de fenómeno mental conciente: el dolor. Las señales de dolor son transmitidas desde las terminales sensoriales nerviosas a la médula espinal por dos tipos de fibras: las fibras A delta se especializan en sensaciones de picazón [*prickling*] y las fibras C se especializan en sensaciones de ardor y malestar [*burning and aching sensations*]. En la médula espinal pasan a través del tracto de Lissauer y terminan en las neuronas de la médula. A medida que las señales suben a la espina dorsal y entran en el cerebro se separan en dos rutas: la ruta de picazón y la ruta de ardor. Ambas rutas pasan a través de una estructura llamada tálamo, pero, más allá de ese nivel, el dolor de picazón [*prickling pain*] se localiza más claramente en el córtex sensorial somático, específicamente en el área somática 1, mientras que la ruta del dolor de ardor [*burning pain*] transmite señales no sólo hacia arriba en el córtex, sino también lateralmente en el hipotálamo y en otras regiones basales del cerebro. Como consecuencia de estas diferencias es mucho más fácil localizar una sensación de picazón; mientras que las sensaciones de ardor y malestar pueden ser más dificultosas porque activan más 'partes' del sistema nervioso. La sensación real de dolor parece ser causada por la estimulación de las regiones basales, especialmente el tálamo, y por la estimulación del córtex sensorial somático.

Ahora bien, a los fines filosóficos, es esencial insistir en este punto: las sensaciones de dolor son causadas por una serie de eventos que comienzan en terminales nerviosas libres y terminan en el tálamo y en otras regiones del cerebro. En lo que concierne a las sensaciones específicas, los eventos en el sistema nervioso central son, por cierto, suficientes para causar dolores, como sabemos por los dolores de miembros amputados [*phantom limb pains*][14] y por los dolores causados por porciones relevantes del cerebro artificialmente estimuladas. Y lo que es verdadero del dolor lo es también de los fenómenos mentales en general. Para decirlo con crudeza, y tomando al resto del sistema nervioso central como parte del cerebro a los fines de esta discusión: todo lo que importa en nuestra vida mental, todos nuestros pensamientos y sentimientos, están causados por procesos dentro del cerebro. En lo que concierne a la causación de los estados mentales, el paso crucial es el que sucede dentro de la cabeza y no el estímulo externo. Y el argumento en favor de esto es, simplemente, que si los eventos fuera del cerebro ocurrieran aunque sin causar nada en el cerebro, no habría eventos men-

14. Los dolores de miembros amputados son dolores que se sienten como proviniendo del ahora miembro inexistente.

tales, mientras que si ocurren eventos en el cerebro, los eventos mentales ocurrirían aun cuando no hubiera estímulo externo.

Creo que estos puntos son obvios, pero son inconsistentes con dos puntos de vista muy comunes acerca de la mente. Uno trata a la causación externa como el modo esencial de causación para los contenidos mentales. Pero en la explicación dada las cadenas causales externas sólo son importantes en la medida en que realmente impactan el sistema nervioso central. Otro punto de vista ampliamente sostenido es que no puede haber una relación causal entre los estados mentales y los cerebrales porque los estados mentales son estados cerebrales, y la forma de la relación de identidad en cuestión excluye la posibilidad de toda relación causal psicofísica. Por ejemplo, muchos filósofos materialistas solían afirmar que los dolores sólo *son* estimulaciones de las fibras C, mientras que según la explicación dada, las estimulaciones de las fibras C no son *idénticas* a los dolores, pero son *parte de las causas* de (ciertos tipos de) dolor.

Formulemos entonces la siguiente pregunta obvia: si los dolores y demás fenómenos mentales son causados por procesos neurofisiológicos, ¿qué son los fenómenos mismos? Bueno, en el caso de los dolores, son obviamente tipos de sensaciones no placenteras, pero esta respuesta nos deja insatisfechos porque no nos aclara, por así decir, cómo localizar los dolores y otros fenómenos mentales, en relación con el resto del mundo en que vivimos. ¿Cómo encajan los dolores en nuestra ontología general? Pienso, nuevamente, que la respuesta a esta pregunta es obvia, aunque tomará algún tiempo exponerla. A nuestra primera afirmación, esto es, que los dolores y demás fenómenos mentales son causados por procesos cerebrales, tenemos que agregar una segunda afirmación: *los dolores y demás fenómenos mentales son características del cerebro.*

Uno de los objetivos primordiales de esta discusión es mostrar cómo esas dos proposiciones pueden ser verdaderas *al mismo tiempo*. Ese par de tesis puede generar diferentes grados de perplejidad filosófica. En un cierto nivel podemos sentirnos perplejos respecto de cómo los fenómenos mentales y los físicos pueden estar en relación causal, cuando uno es una característica del otro. ¿No nos llevaría esto a la espantosa doctrina de la *causa sui*? Esto es, ¿no implicaría que la mente se causa a sí misma? En el fondo, gran parte de nuestra perplejidad proviene de un malentendido acerca de la naturaleza de la causación. Es tentador pensar que siempre que A causa a B tienen que haber dos eventos discretos, uno identificado como la causa y el otro identificado como el efecto; que toda la causación funciona sobre el modelo del relámpago que causa el trueno. Si adoptamos este modelo basto de la causación estaremos tentados de pensar que las relaciones causales entre el cerebro y la mente nos fuerzan a aceptar algún tipo de dualismo; que eventos en un reino, el 'físico', causan eventos en otro reino, el 'mental'. Pero esto me parece un error. Y la manera de resolver el error es adoptar un concepto más sofisticado de causación. Para hacer esto, apartemos por un momento nuestra atención de las relaciones mente-cerebro con el fin de observar otro tipo de relaciones causales que se dan en la naturaleza.

Macro- y micro-propiedades

La distinción entre micro y macro-propiedades de los sistemas, es habitual en la física. Considérese, por ejemplo, el escritorio frente a mí, o el arroyo que fluye fuera de la ventana de mi oficina. Cada sistema está compuesto por micro-partículas y las micro-partículas tienen características en el nivel de las moléculas y de los átomos, así como en el de las partículas subatómicas. Pero cada sistema tiene también ciertas propiedades como la solidez en el caso de la mesa, o la fluidez en el caso del arroyo, que son macro-propiedades o propiedades de superficie [*surface properties*] de los sistemas físicos. Algunas macro-propiedades, pero no todas, pueden ser causalmente explicadas en base al comportamiento de los elementos en el micro-nivel. Por ejemplo, la solidez de la mesa frente a mí es explicada (causalmente) por la estructura reticular [*lattices*] de la que la mesa se compone. De manera similar, la fluidez del agua es explicada (causalmente) por el comportamiento de los movimientos de las moléculas de H_2O. Pero no todas las macro-propiedades tienen una explicación causal en términos de micro-comportamiento. Por ejemplo, la velocidad del arroyo no se explica en base al movimiento de las moléculas sino más bien por el ángulo de la pendiente, la atracción de la gravedad y la fricción provista por el lecho. Pero en el caso de las macro-características que son explicadas causalmente en base al comportamiento de los elementos en el micro-nivel, me parece que tenemos un modelo perfectamente normal para explicar las intrigantes relaciones entre la mente y el cerebro. En el caso de la solidez y la fluidez, no tenemos ninguna dificultad en decir que los fenómenos de superficie son *causados por* el comportamiento de elementos del micro-nivel y, al mismo tiempo, que los fenómenos de superficie *sólo son* rasgos (físicos) de los sistemas en cuestión. Mi modo preferido de enunciar este punto es decir que la característica de superficie F es *cau-*

sada por el comportamiento de los micro-elementos M, y que al mismo tiempo está *realizada en* [*realized in*] el sistema de los micro-elementos. Las relaciones entre F y M son causales pero al mismo tiempo F es, simplemente, una característica de nivel más elevado del sistema que consiste en los elementos M.

En contra de esto se podría decir que F sólo es idéntico a las características de M. Así, por ejemplo, podríamos definir la solidez como la estructura reticular del arreglo molecular [*molecular arrangement*]. Este punto me parece correcto, pero no considero que sea una objeción al análisis que propongo. Es típico del progreso de la ciencia que una expresión que es definida originariamente en términos de las características de superficie de un fenómeno, características accesibles a los sentidos, sea definida subsiguientemente en términos de la micro-estructura que las causa. Así, para tomar el ejemplo de la solidez, la mesa frente a mí es sólida en el sentido ordinario de que es rígida, que resiste cierta presión, que sostiene libros, que no es fácilmente penetrable por otros objetos tales como otras mesas, etcétera. Tal es la noción de sentido común de la solidez. Ahora bien, en vena científica uno puede definir la solidez como cualquier micro-estructura que cause esos toscos rasgos observables. Uno puede decir que la solidez es sólo la estructura reticular del sistema de moléculas y que la solidez así definida causa, por ejemplo, la resistencia al tacto y a la presión; o, uno puede decir que la solidez consiste de cosas tales como la rigidez y la resistencia al tacto y a la presión, y que está causada por el comportamiento de los elementos en el micro-nivel. Este paso de la causación a la identidad definicional es muy común en la historia de la ciencia. Considérese los siguientes pares: un relámpago es causado por una descarga eléctrica —el relámpago sólo es una descarga eléctrica; el color rojo es causado por emisiones de fotones con una longitud de onda de 600 nanómetros— rojo sólo es una emisión de fotones de 600 nanómetros; el calor es causado por movimientos de moléculas —el calor es sólo la energía cinética media de los movimientos de las moléculas.

Si aplicamos estas lecciones al estudio de la mente me parece que no hay dificultad en dar cuenta de las relaciones metafísicas de la mente con el cerebro en términos de una teoría causal del funcionamiento del cerebro como productor de estados mentales. Así como la fluidez del agua está causada por el comportamiento en el micro-nivel y es al mismo tiempo una característica realizada en el sistema de micro-elementos, exactamente en el mismo sentido de 'causado por' y 'realizado en', los fenómenos mentales son causados por los procesos que suceden

en el cerebro en el nivel neuronal o modular, pero están realizados en el mismo sistema que consiste en neuronas organizadas en módulos. Y así como necesitamos la distinción micro-macro para cualquier sistema físico, por las mismas razones necesitamos la distinción micro-macro para el cerebro. Aunque podemos decir de un sistema de partículas que es sólido o líquido, no podemos decir de una partícula dada que sea sólida o líquida. De la misma forma, hasta donde sabemos, aunque podemos decir de un cerebro particular que ese cerebro es conciente o que está experimentando sed o dolor, no podemos decir de una neurona en particular que sienta dolor o que experimente sed. Reiterando el punto una vez más: aunque hay misterios empíricos enormes acerca de cómo funciona en detalle el cerebro, no hay obstáculos lógicos, filosóficos o metafísicos en dar cuenta de la relación entre la mente y el cerebro en términos que nos son completamente familiares respecto del resto de la naturaleza. Nada es más común en la naturaleza que el que los rasgos de superficie de un fenómeno sean causados por una micro-estructura y realizados en ella; y ésas son, exactamente, las relaciones que son exhibidas por la relación de la mente con el cerebro. Las características intrínsecamente *mentales* del universo son las características *físicas* de alto nivel de los cerebros.

La posibilidad de los fenómenos mentales

Retornemos ahora a los cuatro problemas que parece enfrentar toda solución putativa del problema mente-cerebro.

¿Cómo es posible la conciencia?

La mejor forma para mostrar cómo algo es posible es mostrar cómo es en efecto, y ya hemos dado un esquema de cómo los dolores son causados efectivamente por los procesos neurofisiológicos que suceden en el tálamo y en el córtex sensorial. ¿Por qué es que mucha gente no se siente satisfecha con este tipo de respuesta? Creo que si se hace una analogía con un problema anterior en la historia de la ciencia podemos disipar esa sensación de perplejidad. Durante largo tiempo muchos biólogos y filósofos pensaron que era imposible, en principio, dar cuenta de los fenómenos de la vida apelando a un fundamento puramente biológico. Pensaron que además de los procesos biológicos era necesario otro elemento, tenía que postularse algún *élan vital* para dar vida a lo

que de otra manera era materia muerta e inerte. Es muy difícil darse cuenta hoy en día de lo intensa que fue la disputa entre el vitalismo y el mecanicismo, una generación atrás. Hoy esas cuestiones no se toman más en serio. ¿Por qué? ¿Es sólo porque hemos sintetizado la urea (el primer componente orgánico en ser sintetizado) y eso probó que los componentes orgánicos podían ser producidos artificialmente? Pienso que no. Pienso en cambio que es porque hemos llegado a visualizar el carácter biológico de los procesos que son típicos de los organismos vivientes. Una vez que comprendemos cómo las características que son típicas de los seres vivientes tienen una explicación biológica, deja de ser misterioso para nosotros que la materia inerte deba tener vida. Consideraciones exactamente análogas deberían aplicarse a nuestra discusión acerca de la conciencia. Que este trozo de materia inerte, esta sustancia gris y blanca con textura de harina de avena, sea conciente, no debería resultar más misterioso que lo que nos parece problemático que este trozo de materia, esta colección de ácidos nucleicos, proteínas y otras moléculas adosado a una estructura de calcio, esté viva. En síntesis, el modo de disipar el misterio es comprender los procesos. Todavía no entendemos los procesos completamente, pero entendemos el *carácter* de los procesos, entendemos que hay ciertos procesos electroquímicos que acaecen en las relaciones entre neuronas o entre módulos de neuronas y quizás otras características del cerebro, y que esos procesos son causalmente responsables de los fenómenos de conciencia.

¿Cómo pueden tener intencionalidad átomos en el vacío?

Como en el caso de nuestra primera pregunta, la mejor manera de mostrar cómo algo es posible es mostrar cómo es en efecto. Considérese la sed. Hasta donde sabemos, al menos ciertos tipos de sed son causadas en el hipotálamo por secuencias de disparos de neuronas. Estos disparos son a su vez causados por la acción de la hormona peptídica angiotensina II en el hipotálamo, y la angiotensina II, a su vez, es sintetizada por la renina, la cual es secretada por los riñones. La sed, al menos la de estos tipos, es causada por una serie de eventos en el sistema nervioso central, principalmente en el hipotálamo, y se realiza en el hipotálamo. Adviértase que la sed es un estado intencional. Tener sed es tener, entre otras cosas, deseo de beber. La sed tiene contenido proposicional, dirección de ajuste, condiciones de satisfacción y todos los otros rasgos que son comunes a los estados intencionales.

Como con los 'misterios' de la vida y de la conciencia, la manera de aclarar el misterio de la intencionalidad es describir con todo el detalle que podamos cómo los fenómenos son causados por procesos biológicos, al mismo tiempo que se realizan en sistemas biológicos. Las experiencias visuales y auditivas, las sensaciones táctiles, de hambre, de sed, el deseo sexual y las experiencias olfatorias, son todas causadas por procesos cerebrales y se realizan en la estructura del cerebro, y todas son fenómenos intencionales. No estoy diciendo que debamos perder nuestra percepción de los misterios de la naturaleza; por el contrario, los ejemplos que cité son todos, en algún sentido, asombrosos. Pero quiero decir que ellos no son ni más ni menos misteriosos que otros asombrosos rasgos del mundo, como la existencia de la fuerza gravitacional, el proceso de fotosíntesis o el tamaño de la Vía Láctea.

Subjetividad

El enigma de la subjetividad puede enunciarse de manera bastante simple. Desde el siglo diecisiete nuestra concepción de la realidad ha involucrado la noción de objetividad total [*total objectivity*]. La realidad, según esa visión, es aquello que es accesible a todo observador competente. En algunas versiones, la realidad es lo que es medible objetivamente. Ahora bien, la pregunta es, ¿cómo acomodamos la subjetividad de los estados mentales en este cuadro?, ¿cómo encuadramos en una concepción objetiva del mundo real el hecho de que cada uno de nosotros tiene en realidad estados mentales subjetivos? La solución a este enigma también puede enunciarse simplemente. Es un error suponer que la definición de la realidad deba excluir la subjetividad. Si 'ciencia' es el nombre del conjunto de verdades objetivas y sistemáticas que podemos enunciar acerca del mundo, entonces la existencia de la subjetividad es un hecho científico tan objetivo como cualquier otro. Si una explicación científica del mundo intenta describir cómo son las cosas, entonces uno de los rasgos de la explicación será la subjetividad de los estados mentales, ya que es un hecho simple de la evolución biológica el que haya producido ciertos tipos de sistemas biológicos, a saber, el humano y ciertos cerebros de animales, que tienen rasgos subjetivos. Mi estado de conciencia actual es una característica de mi cerebro y en consecuencia me es accesible de una manera que no le es accesible a otro, y el estado de conciencia actual suyo es un rasgo de su cerebro que le es accesible a usted de una manera que no me es accesible a mí. Así, la existencia de la subjetividad es un hecho físico objetivo de la biología. Un error recurrente consiste en tratar de definir 'ciencia' en términos

de ciertas características de las teorías científicas existentes. Pero una vez que ese provincialismo es percibido como el prejuicio no científico que es, entonces cualquier dominio de hechos está sujeto a la investigación científica. Si, por ejemplo, Dios existiera, entonces ese hecho sería un hecho de la ciencia, como cualquier otro. No sé si Dios existe, pero no tengo duda de que los estados mentales subjetivos existen, porque yo tengo ahora uno y usted también. Si el hecho de la subjetividad va en contra de cierta definición de 'ciencia', entonces habría que abandonar la definición y no el hecho.

Causación intencional

Para nuestro propósito, el problema de la causación intencional es cómo dar cuenta de lo mental de modo de evitar el epifenomenalismo. ¿Cómo, por ejemplo, algo tan gaseoso y etéreo como el pensamiento podría dar origen a una acción? La respuesta es que los pensamientos no son gaseosos ni etéreos. Sus propiedades lógicas e intencionales están solidamente fundadas en sus propiedades causales en el cerebro. Los estados mentales pueden causar la conducta mediante el proceso causal ordinario, porque son estados físicos del cerebro. Ellos tienen un nivel elevado y un nivel bajo de descripción, y cada nivel es causalmente real.

Para ilustrar esas relaciones podemos usar nuevamente una analogía con la física. Considérese el martillar un clavo con un martillo. El martillo y el clavo tienen que tener un cierto grado de solidez. Los martillos hechos de algodón o de manteca serían muy poco útiles, y los martillos hechos de agua o de vapor no son martillos. La solidez es una propiedad causal real del martillo y no algo epifenoménico. Pero la solidez misma está causada por el comportamiento de las partículas en el micro-nivel y se realiza en el sistema de los micro-elementos. La existencia de dos niveles causales reales de descripción en el cerebro, un macro-nivel de procesos neurofisiológicos mentales y otro micro-nivel de procesos fisiológicos neuronales, es exactamente análogo a la existencia de los dos niveles causales reales en la descripción del martillo. La conciencia, por ejemplo, es una propiedad causal real del cerebro y no algo epifenoménico. Mi intento conciente de realizar una acción, tal como levantar mi brazo, causa el movimiento del brazo. En un nivel más elevado de descripción, mi intención de levantar el brazo tiene al movimiento de mi brazo como su condición de satisfacción y causa el movimiento del brazo. En un nivel más bajo de descripción, una serie de disparos de neuronas que se originan en el córtex causan la descarga del neurotransmisor acetilcolina en las placas neuromusculares ['*end plates*'] en donde los axones terminales de las neuronas motoras se conectan con las fibras musculares; esto a su vez causa una serie de cambios químicos que resultan en la contracción del músculo. En el caso de martillar un clavo también sucede que la misma secuencia de eventos tiene dos niveles de descripción que son causalmente reales, y el nivel elevado de rasgos causales es causado y realizado en la estructura de los elementos de nivel más bajo.

Categorías tradicionales

Hasta aquí me he resistido a usar el vocabulario tradicional de dualismo, monismo, fisicalismo, etcétera al intentar caracterizar la posición defendida en este capítulo. Sin embargo, puede ser útil ver cómo esos enfoques se relacionan con las categorías tradicionales. En una discusión de estos temas durante la conferencia sobre Filosofía de la Mente que tuvo lugar en la New York University, Hilary Putnam, de Harvard University, caracterizo el enfoque que presento aquí como (1) dualismo de propiedades, (2) emergentismo [*emergentism*], (3) superveniencia [*supervenience*]. Creo que si consideramos cada una de estas evaluaciones profundizaremos la comprensión de estas cuestiones.

Dualismo de propiedades

Si por 'dualismo de propiedades' se entiende, simplemente, el punto de vista de que el mundo contiene algunos rasgos físicos que son mentales —mi actual estado de conciencia, por ejemplo— y algunos rasgos físicos que son no-mentales —el peso de mi cerebro, por ejemplo—, entonces mi punto de vista puede describirse correctamente como un dualismo de propiedades. Creo sin embargo que hay algo profundamente engañoso en esta caracterización. 'Dualismo de propiedades' parece implicar que hay dos y sólo dos tipos de propiedades en el mundo, el físico y el mental, y ése no es de ninguna manera, el punto de vista que sostengo. Para mí, las propiedades mentales sólo son características físicas de alto nivel de ciertos sistemas físicos, en el mismo sentido en que la solidez o la fluidez son características físicas de alto nivel de ciertos sistemas físicos. Así, las propiedades mentales son propiedades físicas en el sentido en que la fluidez y la solidez son propiedades físicas. Me parece que este punto de vista es descripto correctamente no tanto

como dualismo de propiedades sino como polismo de propiedades [*property polyism*]. Esto es, que hay cantidades de tipos diferentes de propiedades de sistemas de alto nivel y que las propiedades mentales están entre ellas. Para decirlo de otro modo, de acuerdo con mi punto de vista las palabras 'mental' y 'físico' no son opuestas entre sí porque las propiedades mentales, interpretadas ingenuamente, sólo son una clase de propiedades físicas, y las propiedades físicas se oponen correctamente no a las propiedades mentales sino a rasgos tales como las propiedades lógicas y las propiedades éticas, por ejemplo.

Emergentismo

Respecto de la pregunta de si debemos o no concebir a las propiedades mentales como emergentes, valen consideraciones similares. Todo depende de lo que uno signifique por 'emergente'. Si vamos a pensar como emergente a una característica de alto nivel del sistema, como la solidez, la fluidez, etcétera, entonces en ese sentido creo que los estados de conciencia, la intencionalidad, la subjetividad, etcétera, son propiedades emergentes de ciertos sistemas biológicos. De hecho, si definimos a las propiedades emergentes de un sistema de elementos como las propiedades que pueden ser explicadas por el comportamiento de elementos individuales pero que no son propiedades de los elementos interpretados individualmente, entonces es una consecuencia trivial de mi punto de vista que las propiedades mentales son propiedades emergentes de los sistemas neurofisiológicos. Sin embargo, tradicionalmente, se ha considerado que emergentismo implica algo misterioso, que hay un proceso misterioso no físico que produce un tipo peculiar de propiedad. En pocas palabras, el emergentismo tiende a compartir los aspectos más misteriosos del dualismo, y en ese sentido niego que mi punto de vista pueda caracterizarse correctamente como emergentista. Si se considera que el emergentismo implica algo misterioso en la existencia de las propiedades emergentes, algo que yace más allá del alcance de las ciencias físicas o biológicas tal como son normalmente interpretadas, entonces me parece claro que las propiedades mentales no son emergentes en ese sentido.

Superveniencia

La doctrina de la superveniencia de lo mental en lo físico dice que no puede haber diferencias mentales sin las correspondientes diferencias

físicas: si en dos momentos diferentes un sistema está en dos estados mentales diferentes, entonces en esos dos momentos tiene que tener propiedades físicas diferentes. Este enfoque es una consecuencia de la tesis de que los fenómenos mentales son causados por el cerebro y realizados en él, porque si los efectos son diferentes, las causas tienen que ser diferentes. Me parece, por cierto, que es un mérito del enfoque que he adelantado aquí, que la superveniencia de lo mental sea simplemente un caso especial del principio general de la superveniencia de macro-propiedades en micro-propiedades. No hay nada especial o arbitrario o misterioso acerca de la superveniencia de lo mental en lo físico; es simplemente un caso más de la superveniencia de las propiedades físicas de alto nivel en las propiedades físicas de nivel más bajo. Si un recipiente con agua tiene hielo en cierto momento y líquido en otro momento, entonces tiene que haber una diferencia en el comportamiento de las micro-partículas que dé cuenta de la diferencia. De manera semejante, si yo quiero agua en un momento y luego no quiero agua, tiene que haber una diferencia en mi cerebro que dé cuenta de esta diferencia en mis estados mentales.

Consecuencias para la filosofía de la mente

Algunos conceptos mentales como, por ejemplo, *tener un dolor* o *creer* que tal y cual, denotan entidades que existen enteramente en la mente. Otras como *ver* o *conocer* también se refieren a fenómenos mentales pero requieren que se satisfagan condiciones adicionales para que el concepto sea aplicable. Así, por ejemplo, decir que X sabe que P implica algo más que X crea que P; implica, entre otras cosas, que P es verdadero, y que la verdad de P no puede ser, en general, algo que suceda únicamente en la mente de X. Decir que X ve P implica que X tiene una experiencia visual de cierto tipo, pero también implica que es el caso que P. Llamemos a los conceptos cuyas condiciones de verdad dependen sólo de lo que sucede en la mente 'conceptos mentales puros' ['*pure mental concepts*'], y llamemos a los conceptos mentales cuyas condiciones de verdad requieren fenómenos extra-mentales, 'conceptos mentales híbridos' ['*hybrid mental concepts*']. Ahora bien, dado que los conceptos mentales híbridos contienen por definición un componente mental, en la medida en que discutimos la naturaleza de la mente, podemos separar el componente mental y examinarlo separadamente. Para cada concepto mental híbrido hay un concepto mental puro que le

corresponde, que capta el componente mental puro del concepto híbrido. En lo que concierne a la mente, podemos confinar nuestra discusión a los conceptos mentales puros y a los estados mentales puros que son las denotaciones [*denotations*] de los conceptos mentales puros. Toda vez que un fenómeno mental está presente en la mente de un agente —por ejemplo, siente dolor, piensa en la filosofía o desea beber una cerveza fría— las condiciones causalmente suficientes para el fenómeno están enteramente en el cerebro. Y por cierto que la tesis de que los fenómenos mentales están causados por el cerebro y realizados en él tiene la consecuencia de que, para cualquier fenómeno mental, hay condiciones causalmente suficientes en el cerebro. Llamenos a este principio, *el principio de suficiencia neurofisiológica* [*the principle of neurophysiological sufficiency*]. Ahora bien, si este principio es verdadero, entonces muchas teorías corrientes en la filosofía de la mente se tornarán falsas, porque son inconsistentes con él. Por ejemplo, varios filósofos que siguen a Wittgenstein y a Heidegger han tratado de explicar la intencionalidad de los fenómenos mentales en términos de relaciones sociales. Pero, ¿cómo hemos de tomar esa explicación? Si la tomamos como afirmando que las relaciones sociales son necesarias para la vida mental o constitutivas de ella, entonces sabemos que tiene que ser falsa, porque las relaciones sociales son relevantes para la producción causal de la intencionalidad sólo si impactan en los cerebros de los agentes humanos; y los estados mentales efectivos, las creencias, los deseos, las esperanzas, los temores y el resto de ellos, tienen condiciones causalmente suficientes que son internas, enteramente, al sistema nervioso. Esto no implica negar que las relaciones sociales sean cruciales para la producción de muchas formas de intencionalidad, tal como, por ejemplo, el lenguaje. Los niños pueden aprender y usar un lenguaje sólo si están expuestos a otras personas que también usan lenguaje. Pero la tesis de que hay formas de intencionalidad que requieren una base social necesita ser reinterpretada de modo de que sea consistente con la afirmación de que la intencionalidad es un producto puramente interno de los procesos fisiológicos internos. Esos enfoques no son necesariamente inconsistentes; sólo pueden interpretarse como formas de describir aspectos diferentes del mismo fenómeno. El error consiste en suponer que las relaciones sociales puedan de una manera u otra reemplazar o substituir lo que sucede en el cerebro.

Una negación implícita más destacada del principio de suficiencia neurofisiológica proviene de la tradición construida en torno a la afirmación de Wittgenstein de que 'un proceso interno requiere un criterio externo'.[15] Así, por ejemplo, Norman Malcolm ha tratado de dar un explicación no interna del soñar,[16] Elizabeth Anscombe ha tratado de explicar las intenciones en términos de conducta externa,[17] y Anthony Kenny ha tratado de explicar muchas emociones en términos de su escenario social [*social setting*] y de sus consecuencias conductuales.[18] Pero es difícil interpretar esos análisis de modo que sean consistentes con el principio de suficiencia neurofisiológica. Cualesquiera sean los demás rasgos que los sueños puedan poseer, son causados por procesos neurofisiológicos. Y lo mismo vale para las intenciones y las emociones, como el miedo y la angustia. Ahora bien, quizá podríamos interpretar los enfoques de Malcolm, de Anscombe y de Kenny como describiendo simplemente restricciones a la posesión de un *vocabulario* para discutir los fenómenos mentales. Y quizá podríamos interpretar la aseveración de Wittgenstein como la aseveración de que un *vocabulario* que corresponda a los procesos internos requiere criterios externos. Pero si tomamos esas aseveraciones como aseveraciones acerca de la *naturaleza* de los fenómenos mentales mismos —esto es, que uno no puede tener un sueño o una intención o estar enojado a menos que ciertas condiciones externas sean satisfechas, condiciones externas al cerebro—, entonces sabemos que esas tesis tienen que ser falsas debido al principio de suficiencia neurofisiológica. Lo que sucede en la cabeza tiene que ser causalmente suficiente para cualquier estado mental.

Y, por supuesto, la tradición wittgensteiniana es en sí misma parte de una tradición mayor que persigue un análisis conductista o cuasi-conductista de los conceptos mentales. Y una vez más, por el principio de suficiencia neurofisiológica sabemos que esos esfuerzos están condenados al fracaso. No podemos definir los fenómenos mentales en términos de sus manifestaciones conductuales, porque sabemos que siempre es posible experimentar los fenómenos, con independencia de tener alguna manifestación conductual.

15. Wittgenstein, Ludwig (1973): *Philosophical Investigations*, trad. G.E.M. Anscombe, Nueva York, Macmillan.

16. Malcolm, Norman (1959): *Dreaming*, Londres, Routledge & Kegan Paul.

17. Anscombe, G.E.M. (1963): *Intention*, Ithaca, Nueva York, Cornell Univ. Press.

18. Kenny, Anthony (1963): *Action, Emotion and Will*, Londres, Routledge & Kegan Paul.

Algunas conclusiones

Los objetivos más polémicos de este capítulo respecto de las mentes y los programas, pueden sintetizarse con rapidez. Para que la claridad sea total, enunciaré un conjunto de 'axiomas' y derivaré las conclusiones relevantes.

Axioma 1. Los cerebros causan a las mentes.

Ésta es, sencillamente, la enunciación cruda del hecho empírico de que procesos causales relevantes en el cerebro son suficientes para producir cualquier fenómeno mental. Es importante volver a enfatizar que en lo que hace a los fenómenos mentales puros, no hay ninguna conexión esencial entre los procesos causales internos que son suficientes para los fenómenos mentales y las relaciones causales de *input-output* del sistema total. En principio, podríamos vivir toda nuestra vida mental sin tener ninguno de los estímulos apropiados o ningún comportamiento externo normal.

Axioma 2. La sintaxis no es suficiente para la semántica.

Ésta es una verdad conceptual o lógica que articula la distinción entre el nivel de los símbolos formales y el nivel del significado.

Axioma 3. Las mentes tienen contenidos; específicamente, tienen contenidos intencionales o semánticos.

Axioma 4. Los programas son definidos formalmente o sintácticamente.

Ahora bien, de estos puntos obvios podemos derivar algunas conclusiones discutibles.

Conclusión 1. En sí misma, la instanciación de un programa nunca es suficiente para tener una mente (por los axiomas 2, 3 y 4).

Esta conclusión es suficiente por sí misma para refutar a la inteligencia artificial fuerte.

Conclusión 2. La manera en que el cerebro causa a las mentes no puede ser sólo por la instanciación de un programa (axioma 1 y conclusión 1).

Conclusión 3. Cualquier artefacto que tenga una mente tendría que tener poderes causales equivalentes (al menos) a los del cerebro (por axioma 1 trivialmente).

Conclusión 4. Para cualquier artefacto que tenga una mente, el programa por sí mismo no sería suficiente para proveerle tal mente. El artefacto tendría que tener poderes causales equivalentes al cerebro (por las conclusiones 1 y 3).

Quienquiera que desee cuestionar las tesis centrales nos debe una especificación precisa de los 'axiomas' y las derivaciones que cuestione.

TRADUCTORA: Florencia Luna.
REVISIÓN TÉCNICA: Eduardo Rabossi.