

# *¿Puede la neurobiología enseñarnos algo sobre la conciencia?1 \**

PATRICIA SMITH CHURCHLAND  
*Universidad de California, San Diego*

## I. INTRODUCCIÓN

Los sistemas nerviosos humanos exhiben una lista impresionante de capacidades complejas, incluyendo las siguientes: percibir, aprender y recordar, planear, decidir, ejecutar acciones, así como las capacidades de estar despierto, dormirse, soñar, prestar atención, y darse cuenta. Aunque la neurociencia ha avanzado espectacularmente en este siglo, todavía no comprendemos con detalles satisfactorios cómo cualquier capacidad de la lista emerge de las redes de neuronas<sup>2</sup>. No alcanzamos a comprender de manera completa cómo los humanos pueden ser conscientes, pero tampoco comprendemos cómo pueden pasear, correr, subir a los árboles o saltar con pértiga. También, cuando uno se aparta de todo eso, la conciencia es intrínsecamente más misteriosa que el control motor. Equilibrado con la decepción de que una plena comprensión todavía nos esquiva, tenemos el optimismo prudente, basado principalmente en la naturaleza del progreso tras nosotros. Pues la neurociencia cognitiva ya ha sobrepasado ampliamente lo que los filósofos escépticos consideraron antaño posible, y un progreso continuo parece probable.

\* Traducción de Pascual F. Martínez-Freire

1 Este artículo se basa en mi Discurso Presidencial a la Asociación Filosófica Americana, División Pacífico, en marzo de 1993, publicado en *Proceedings and Addresses of the APA* (1994). Estoy en deuda con mucha gente por su discusión y crítica: Francis Crick, Rodolfo Llinás, Antonio Damasio, Hanna Damasio, Joe Bogen, Ramachandran, Robert Van Gulick, Owen Flanagan, Dan Lloyd, y especialmente Paul Churchland.

2 Véase nuestra discusión en Churchland y Sejnowski (1992), *The Computational Brain*.

Al asumir que la neurociencia puede revelar los mecanismos físicos que hacen posible las funciones psicológicas, estoy asumiendo que es en verdad el cerebro el que ejecuta esas funciones, que las capacidades de la mente de los humanos son de hecho capacidades del cerebro humano. Esta asunción, y su rechazo anejo de las almas cartesianas o espíritus o «estofa fantasmal» existente separadamente del cerebro, no es una fantasía. Al contrario, es una hipótesis altamente probable, basada en pruebas disponibles actualmente de la física, la química, la neurociencia y la biología evolucionista. Al decir que el fisicalismo es una hipótesis, quiero destacar su condición de asunto empírico. No asumo que es una cuestión de análisis conceptual, una intuición *a priori*, o una fé religiosa, aunque comprendo que no todos los filósofos están de acuerdo conmigo en este punto<sup>3</sup>.

De modo adicional, estoy convencida de que la estrategia correcta para comprender las capacidades psicológicas es esencialmente reduccionista, por lo que entiendo, en general, que comprender los mecanismos neurobiológicos no es un adorno sino una necesidad. Si la ciencia tendrá éxito finalmente en reducir los fenómenos psicológicos a fenómenos neurobiológicos es, no hay que decirlo, también otra cuestión empírica. Adoptar la estrategia reduccionista quiere decir intentar explicar los niveles macro (propiedades psicológicas) en términos de niveles micro (propiedades de redes neuronales).

La razón fundamental detrás de esta estrategia de investigación es sencilla: si quieres comprender cómo funciona una cosa, necesitas comprender no sólo su perfil conductual sino también sus componentes básicos y cómo se organizan para constituir un sistema. Si tú no tienes los diseños de ingeniería disponibles para el caso, recurres a la ingeniería inversa - la táctica de tomar una parte de un mecanismo para ver cómo funciona<sup>4</sup>. En la medida en que estoy intentando descubrir explicaciones macro-a-micro, soy una reduccionista. Puesto que muchos filósofos que están de acuerdo conmigo sobre la naturaleza del alma basada en el cerebro sin embargo despotrican contra el reduccionismo como ridículo si no enteramente despreciable, puede incumbirme comenzar explicando brevemente lo que entiendo y, más importante, lo que *no* entiendo por una estrategia de investigación reduccionista<sup>5</sup>.

3 Para opiniones que están de acuerdo, véase también Francis Crick (1994), Paul Churchland (1989), Daniel Dennett (1991), Owen Flanagan (1992), William G. Lycan (1987), John Searle (1993).

4 Tal como P. S. Churchland y T. J. Sejnowski arguyen en (1989).

5 Para una discusión destacada del reduccionismo que incluye muchas de las complejidades de las que no me ocupo apenas aquí, véase Schaffner (1993).

Aclarando lo negativo primero, puedo decir que *no* entiendo que una estrategia de investigación reduccionista implique que deba adoptarse una *estrategia puramente de abajo-arriba*. Hasta donde puedo decirlo, nadie en neurociencia piensa que el modo de comprender el sistema nervioso sea comprender primero todo acerca de las moléculas básicas, después todo acerca de cada neurona y cada sinapsis, y continuar lentamente así para ascender los diversos niveles de organización hasta que, por fin, uno llega al nivel superior –los procesos psicológicos (Figura 1). Tampoco hay nada en la historia de la ciencia que diga que una estrategia de investigación es reduccionista sólo si es puramente de abajo-arriba. Tal caracterización es paja completamente. La investigación detrás de los éxitos reduccionistas clásicos– explicación de la termodinámica en términos de mecánica estadística, de la óptica en términos de radiación electromagnética, de la transmisión hereditaria en términos de ADN –ciertamente no se ajustaron a ninguna directiva de investigación puramente de abajo– arriba.

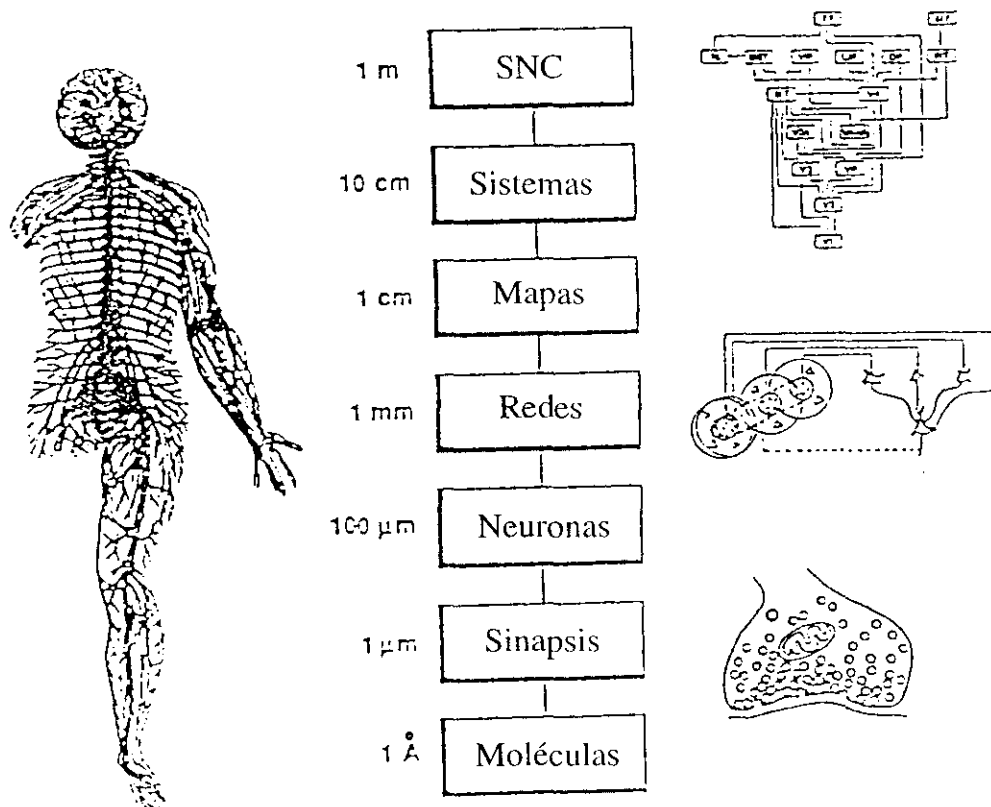


Figura 1

Ilustración esquemática de los niveles de organización en el sistema nervioso. Las escalas espaciales en las que pueden identificarse las organizaciones anatómicas

varían en muchos órdenes de magnitud. El dibujo a la izquierda presenta el «hombre neuronal», mostrando el cerebro, la médula espinal y los nervios periféricos. Los dibujos a la derecha representan estructuras en distintos niveles: (arriba) un subconjunto de áreas visuales en el córtex visual, (en el medio) un modelo de red proponiendo cómo las células ganglionares pueden conectarse con células «simples» en el córtex visual, y (abajo) una sinapsis química. (Tomado de Churchland y Sejnowski 1992).

En lo que concierne a la neurociencia y a la psicología, mi punto de vista es simplemente que sería lo más acertado conducir la investigación sobre muchos niveles simultáneamente, desde lo molecular, a través de redes, sistemas, áreas cerebrales, y por supuesto conducta. Aquí, como en cualquier otro lugar de la ciencia, las hipótesis en varios niveles pueden *co-evolucionar* en cuanto se corrigen e informan entre sí. Los neurocientíficos serían tontos si se empeñaran en ignorar los datos psicológicos, por ejemplo, al igual que los psicólogos serían tontos si se empeñaran en ignorar todos los datos neurobiológicos.

En segundo lugar, por estrategia de investigación reduccionista no entiendo que haya algo vergonzoso, no científico o indeseable de otra manera en las descripciones o capacidades de alto nivel *per se*. Parece enteramente obvio, por tomar un ejemplo simple, que en los sistemas nerviosos ciertas propiedades rítmicas son propiedades de red que resultan de los rasgos de membrana individual de varios tipos de neurona en la red, junto con el modo en que el conjunto de neuronas interactúan. El reconocimiento de que algo es el rostro de Arafat, por poner otro ejemplo, emerge casi ciertamente de los perfiles de reacción de las neuronas en la red más los modos en que estas neuronas interactúan. «Emergencia» en este contexto está significando, de una manera no fantasmal y respetable, en una primera aproximación, «propiedad de la red». Determinar de manera precisa lo que es la propiedad de red, para alguna proeza particular, supondrá naturalmente un montón de esfuerzo experimental. Además, dado que la conducta neuronal es altamente no lineal, las propiedades de red *nunca* son una simple «suma de partes». Son alguna función - alguna *complicada* función - de las propiedades de las partes. Las capacidades de alto nivel claramente existen, y las descripciones de alto nivel se necesitan por tanto para especificarlas.

El «materialismo eliminativo» se refiere a la hipótesis de que (1) el materialismo es con la mayor probabilidad verdadero, y también (2)

muchos aspectos tradicionales de la explicación de la conducta humana son probablemente no adecuados para la realidad de la etiología de la conducta<sup>7</sup>. La analogía estándar aquí es que al igual que el «fluido calórico» fué útil pero fundamentalmente erróneo para comprender los fenómenos térmicos (conducción, convección, radiación) también algunas categorías psicológicas actualmente invocadas pueden ser de algún modo útiles pero fundamentalmente erróneas para desentrañar la etiología conductual. Otras caracterizaciones existentes de las capacidades pueden tener un núcleo de adecuación pero pueden sufrir retoques importantes, análogamente a como la noción de Mendel de «factor» llegó a modificarse por la genética en la noción de «gen» que a su vez fué modificada y profundizada con el desarrollo de la biología molecular. Algunas categorías tales como «actitud» son extremadamente vagas y podrían reemplazarse enteramente; otras, tales como «está durmiendo» ya han sufrido una descomposición conforme la investigación neurofisiológica y electroencefalográfica ha revelado diferencias cerebrales importantes en los varios estadios del sueño. Categorías tales como «memoria», «atención» y «razonamiento» están sufriendo igualmente una revisión, conforme la psicología experimental y la neurociencia avanzan<sup>8</sup>. Queda por ver si hay una realidad neurobiológica para sostener nociones tales como «creencia» y «deseo» tal como han sido articuladas por filósofos actuales como Fodor<sup>9</sup> y Searle<sup>10</sup>, aunque Paul Churchland y yo hemos argumentado que la revisión aquí también es lo más probable<sup>11</sup>. La predicción de revisión también es una hipótesis empírica, y para ella ya existe apoyo empírico<sup>12</sup>.

La posibilidad de revisión no trivial e incluso de reemplazo de descripciones de alto nivel existentes por categorías de bajo nivel «neurobiológicamente armoniosas» es el punto básico que hace *eliminativo* al materialismo eliminativo<sup>13</sup>. Entiendo por categorías «neurobiológicamente armoniosas» aquellas que permiten explicaciones coherentes e

7 Véase la caracterización y defensa de este punto de vista en Paul Churchland (1981), reimpresso en P. M. Churchland (1989).

8 Véase P. S. Churchland, *Neurophilosophy* (1986).

9 Jerry Fodor (1990).

10 John Searle (1992).

11 Op. Cit.

12 Véase Churchland y Sejnowski (1992), Paul M. Churchland (1989).

13 O bien, tal como hemos preferido pero decidido no decir, «lo que hace *revisionista* al materialismo revisionista» (P. S. Churchland 1987). Véase también P. M. Churchland (1993). Para una descripción relacionada aunque algo diferente, véase Bickle (1992).

integradas desde el cerebro entero hacia abajo a través de sistemas neuronales, grandes redes, microredes y neuronas. Sólo un hombre inventado es tan tonto como para pretender que no hay capacidades de alto nivel, que no hay fenómenos de alto nivel<sup>14</sup>. En su aspecto general, mi tesis aquí refleja simplemente este hecho: en un sentido profundamente importante no comprendemos exactamente lo que, en sus niveles superiores, hace realmente el cerebro. Según ello, es práctico reservar incluso nuestras intuiciones más queridas acerca de la función mente/cerebro como hipótesis revisables en vez de como certezas absolutas transcendentales o dadas introspectivamente. El reconocimiento de tal revisabilidad marca una enorme diferencia en cómo conducimos experimentos psicológicos y neurobiológicos, y en cómo interpretamos los resultados.

## II. RECHAZANDO EL OBJETIVO NEUROBIOLÓGICO

Durante las últimas décadas, varios filósofos han expresado reservas respecto del objetivo de investigación reduccionista de descubrir los mecanismos neurobiológicos de las capacidades psicológicas, incluyendo la capacidad de ser consciente. Por tanto, puede ser útil considerar la base de algunas de estas reservas para determinar si justifican abandonar el objetivo, o si deberían desalentar nuestras esperanzas acerca de lo que podría descubrirse sobre la mente/cerebro. Consideraré aquí cinco clases especiales de objeción. Como una concesión a la brevedad, mis respuestas serán implacablemente sucintas, sacrificando los detalles en beneficio de lo más esencial.

### A. El objetivo es absurdo (incoherente)

Un conjunto de razones para condenar la estrategia de investigación reduccionista se resume así: «Simplemente no puedo imaginar que ver azul o el sentimiento de dolor, por ejemplo, puedan consistir en algún patrón de actividad de las neuronas en el cerebro», o, más francamente, «no puedo imaginar cómo puedes sacar conciencia de la carne». Hay a veces una considerable diferencia entre la premisa «es inimaginable» y la conclusión «es imposible», pero hasta donde llego tal diferencia típicamente encubre el núcleo falaz del argumento<sup>15</sup>.

14 Ibid. Véase también P. M. Churchland y P. S. Churchland (1990).

15 Por ejemplo, Colin McGinn (1990).

Dado el escaso detalle con que actualmente comprendemos cómo el cerebro humano «neuronea» cualquiera de sus diversas capacidades, es enteramente predecible que debemos tener dificultad en imaginar los mecanismos neuronales. Cuando la comunidad científica humana era comparablemente ignorante de temas tales como valencia, capas de electrones, etc., los filósofos naturales no podían imaginar cómo se podía explicar la maleabilidad de los metales, la magnetizabilidad del hierro y la resistencia a la oxidación del oro, en términos de componentes subyacentes y de su organización. Hasta la llegada de la biología molecular, mucha gente pensaba que era inimaginable, y por tanto imposible, que ser una cosa viva podía consistir en una organización particular de moléculas «muertas». «No puedo imaginar cómo se puede obtener *vida* a partir de material *muerto*», decían los vitalistas.

Desde la posición ventajosa de la ignorancia considerable, el fracaso para imaginar alguna posibilidad es sólo eso: un fracaso de imaginación- una capacidad psicológica entre otras. No anuncia ningunas limitaciones metafísicas sobre lo que podemos llegar a comprender, y no puede predecir algo significativo acerca del futuro de la investigación científica. Después de reflexionar sobre la complejidad respetable del problema de la termoregulación en homeotermos tales como nosotros, encuentro que no puedo imaginar cómo los cerebros controlan la temperatura corporal bajo condiciones diversas. Sospecho, sin embargo, que esto es un hecho psicológico relativamente poco interesante acerca de mí, que refleja meramente mi actual estado de ignorancia. No es un hecho metafísico interesante acerca del universo ni siquiera un hecho epistemológico acerca de los límites del conocimiento científico.

Una variación de la propuesta «no puedo imaginar» se expresa como «nunca, nunca podemos saber...», o «es imposible comprender alguna vez...» o «está para siempre más allá de la ciencia mostrar que...». La idea aquí es que la imposibilidad de concebir algo dice algo decisivo sobre su imposibilidad empírica o lógica. No insisto en que tales propuestas nunca son relevantes. A veces pueden serlo. Pero son sorprendentemente arbitrarias cuando la ciencia está en los más primitivos estadios del estudio de un fenómeno.

El punto central aquí es que variadas «certezas a priori», en el curso de la historia, han resultado ser falsedades empíricas, aunque obvias y sinceras en sus buenos tiempos. La imposibilidad de que el espacio sea no-euclídeo, la imposibilidad de que en el espacio real líneas paralelas convergieran, la imposibilidad de tener buenas pruebas de que algunos eventos sean indeterminados, o que alguien esté ahora soñando, o que el universo tuvo un principio- cada una perdió su atadura lógica conforme

llegamos a una más profunda comprensión de cómo son las cosas. Si hemos aprendido algo de los muchos descubrimientos contraintuitivos en la ciencia es que nuestras intuiciones pueden ser erróneas. Nuestras intuiciones acerca de nosotros mismos y de cómo funcionamos pueden también ser completamente erróneas. No hay ninguna base en la teoría evolucionista, en la matemática, o en algún otro lugar, para asumir que las concepciones precientíficas son concepciones adecuadas esencial y científicamente.

Una tercera variación de este tema «no, no, nunca» extrae conclusiones acerca de cómo el *mundo tiene que ser realmente*, basadas en *propiedades lingüísticas* de ciertas categorías centrales del uso actual para describir el mundo. Permítaseme dar un ejemplo sencillo: «La categoría ‘mental’ está alejada en el significado- quiere decir algo completamente diferente- de la categoría ‘físico’. Es absurdo por tanto hablar del cerebro viendo o sintiendo, al igual que es absurdo hablar de la mente teniendo neurotransmisores o conduciendo corriente». Supuestamente, esta absurdidad categorial suprime la auténtica posibilidad de que la ciencia pueda descubrir que sentir dolor es actividad en las neuronas del cerebro. El epíteto «error categorial» es a veces considerado suficiente para revelar el sinsentido desnudo del reduccionismo.

Mucho ya ha sido dicho sobre este asunto en otros lugares<sup>16</sup>, y yo evitaré una larga discusión de filosofía del lenguaje mediante tres breves puntos. (1) Es más bien inverosímil suponer que las intuiciones en filosofía del lenguaje pueden ser una guía fiable de lo que la ciencia puede y no puede descubrir acerca de la naturaleza del universo. (2) Los significados *cambian* conforme la ciencia hace descubrimientos acerca de lo que *es* algún macrofenómeno en términos de su composición y de la dinámica de su estructura subyacente. (3) Es improbable que los científicos detengan su investigación si se les informa de que sus hipótesis y teorías «parecen raras» en relación al uso corriente. Más probablemente dirán: «Las teorías podrían parecerle raras, pero déjame enseñar la ciencia de transfondo que nos hace pensar que la teoría es verdadera. Entonces parecerá menos rara». Puede observarse que parecía raro a los contemporáneos de Copérnico decir que la Tierra es un planeta y se mueve; parecía raro decir que el calor es movimiento molecular y que el espacio físico es no-euclídeo y que no hay ningún «abajo» absoluto». Y así sucesivamente.

Que una teoría científicamente plausible parezca rara es sólo un criterio de que no se ha convertido en moneda común, no de que sea erró-

16 Véase por ejemplo Feyerabend (1981).



nea. Los descubrimientos científicos de que cierto macrofenómeno es un resultado complejo de la estructura micro y de su dinámica son típicamente sorprendentes y típicamente parecen raros- al principio. Obviamente nada de esto es una prueba positiva de que podemos realizar una reducción de los fenómenos psicológicos a los fenómenos neurobiológicos. Sólo dice que parecer raro no significa nada, de un modo u otro.

B. El objetivo es inconsistente con la «realizabilidad múltiple»

El núcleo de esta objeción es que si un macrofenómeno puede ser el resultado de más de un mecanismo (organización y dinámica de componentes), entonces no puede identificarse con algún mecanismo, y por ende la reducción del macrofenómeno *al* (singular) microfenómeno subyacente es imposible. Esta objeción me parece totalmente sin interés para la ciencia. De nuevo, permítaseme ignorar detalles importantes y meramente resumir la principal idea clave de las réplicas.

(1) Las explicaciones, y por tanto las reducciones, son relativas a un dominio. En biología, puede ser fructífero detallar primero los principios generales que explican algún fenómeno visto en diversas especies, y luego resolver cómo explicar las diferencias interespecíficas, y luego, si es deseable, cómo explicar diferencias entre individuos dentro de una especie dada. Así los principios generales de cómo funcionan corazones o estómagos son resueltos quizás basados en estudios de una especie única, y las particularidades pueden ser resueltas después. Corazones de rana, corazones de macaco y corazones humanos funcionan de la misma manera general esencialmente, pero hay también diferencias significativas, aparte del tamaño, que exigen análisis comparativos. Consideremos otros ejemplos: (a) a partir de la solución general al problema de copiado que surgió del descubrimiento de la estructura fundamental del ADN, fué posible emprender exploraciones de cómo diferencias en el ADN podían explicar ciertas diferencias en el fenotipo; (b) a partir de la solución general al problema de cómo las neuronas envían y reciben señales, fué posible emprender una exploración detallada sobre las diferencias en perfiles de reacción de clases distintas de neuronas<sup>17</sup>.

(2) Una vez que los mecanismos de algunos procesos biológicos han sido descubiertos, puede ser posible inventar dispositivos para imitar esos procesos. Sin embargo, la invención de la tecnología para corazones artificiales o riñones artificiales no elimina el progreso explicativo sobre corazones reales y riñones reales; no contradice el logro

17 Véase también Owen Flanagan (1996).

reductivo. De nuevo, la posibilidad de que el material hereditario de un tipo diferente de ADN pudiera encontrarse en las cosas en otro lugar del universo no afecta al andamiaje básico de una reducción en este planeta. La ciencia habría sido mucho más pobre si Crick y Watson hubieran abandonado su proyecto a causa de la abstracta posibilidad de material hereditario marciano o material hereditario artificial. De hecho, conocemos ciertamente lo esencial del mecanismo de copiado *en la Tierra* - a saber, el ADN, y conocemos ciertamente mucho sobre cómo cumple su tarea. De manera similar, la ingeniería de neuronas artificiales y de redes neuronales artificiales (RNA) facilita y es facilitada mediante aproximaciones neurobiológicas sobre cómo funcionan las neuronas reales; las tareas ingenieriles no quieren decir que la búsqueda de los principios básicos de la función del sistema nervioso esté equivocada.

(3) Siempre hay cuestiones que quedan para ser contestadas en la ciencia, y por ello llegar a captar el funcionamiento general de un mecanismo, tal como el descubrimiento del emparejamiento de bases en el ADN, no debe confundirse con el ideal utópico de una reducción completa - una completa explicación. Los descubrimientos acerca del funcionamiento general de algo plantean típicamente multitud de cuestiones acerca de su *detallado* funcionamiento, y luego acerca de los detalles de los *detalles*. Para señalar la incompletud de las explicaciones, quizás debíamos evitar la expresión «reducción» en favor de «contacto reductivo». Por tanto debíamos decir que el objetivo de la neurociencia es hacer un contacto reductivo rico con la psicología conforme las dos amplias disciplinas co-evolucionan. Yo misma he experimentado con esta recomendación, y aunque algunos filósofos se entusiasman con ella, los científicos la encuentran curiosamente pedante. En cualquier caso, el «contacto reductivo» entre la biología molecular y la macrobiología se ha hecho firmemente más rico desde 1953, aunque permanecen muchas cuestiones. El contacto reductivo entre psicología y neurociencia se ha hecho también más rico, especialmente en la última década, aunque es justo decir que en general los principios básicos de cómo funciona el cerebro se comprenden escasamente.

(4) ¿Cuáles, precisamente, se supone que son las secuelas programáticas del argumento de realizabilidad múltiple? ¿Es que la neurociencia es *irrelevante* para la comprensión de la naturaleza de la mente humana? Obviamente no. ¿Que la neurociencia *no es necesaria* para comprender la mente humana? No se puede, ciertamente, negar que es notablemente útil. Consideremos los descubrimientos relativos a dormir, estar despierto y soñar; los descubrimientos relativos a cerebros divididos, humanos con lesiones cerebrales focales, la neurofisiología y

neuroanatomía del sistema visual, etc. ¿Quizás no deberíamos mantener nuestras esperanzas demasiado altas? ¿Qué es, en forma precisa, «demasiado altas» aquí? ¿Es la esperanza de que descubriremos los principios generales de cómo funciona el cerebro? ¿Por qué es esa una esperanza demasiado alta?

### C. El cerebro *causa* conciencia

Rechazar el objetivo reduccionista y al tiempo mantener el dualismo a distancia es una maniobra que requiere gran delicadeza. La estrategia de John Searle es decir que aunque el cerebro *causa* estados de conciencia, cualquier identificación de estados conscientes con actividades cerebrales es incorrecta. Tradicionalmente, se ha opinado que lo más que el reduccionista puede esperar son *correlaciones* entre estados subjetivos y estados cerebrales, y aunque las correlaciones pueden ser prueba de causalidad no son prueba de identidad. Searle ha intentado reforzar esa objeción diciendo que mientras que las identificaciones *a / b* en otros lugares de la ciencia revelan la realidad detrás de la apariencia, en el caso de la conciencia la realidad y la apariencia son inseparables - no hay realidad para la conciencia excepto lo que está presente en la conciencia. No hay, por tanto, ninguna reducción a realizar.

Sinópticamente, he aquí por qué la maniobra de Searle es inconvincente: fracasa en apreciar por qué los científicos optan por identificaciones cuando lo hacen. Dependiendo de los datos, las identificaciones entre niveles al efecto de que *a* es *b* pueden ser menos dificultosas y más completas científicamente que suponer que la cosa *a* causa la cosa separada *b*. Esto se vé mejor con ejemplos<sup>18</sup>.

La ciencia, tal como sabemos, dice que la corriente eléctrica en un alambre no está causada por electrones en movimiento; *es* electrones en movimiento. Los genes no están causados por trozos de pares de bases en el ADN; *son* trozos de pares de bases (aunque a veces trozos distribuidos). La temperatura no está causada por energía cinética molecular media; *es* energía cinética molecular media. Reflexionemos por un momento sobre la inventiva requerida para generar explicaciones que mantienen la *no identidad* y causal dependencia de (a) corriente eléctrica y electrones en movimiento, (b) genes y trozos de ADN, y (c) calor y movimiento molecular. No familiarizado con los datos convergentes re-

<sup>18</sup> En la discusión siguiente las ideas se deben sobre todo a Paul Churchland (1994). Para tal discusión véase «Betty Crocker's Theory of the Mind: A Review of John Searle's *The Rediscovery of the Mind*». *The London Review of Books*.

levantes y los éxitos explicativos, uno puede suponer que esto no es tan difícil. Entra Betty Crocker.

En su libro de cocina de horno de microondas, Betty Crocker ofrece explicar cómo funciona un horno de microondas. Dice que cuando enciendes el horno, las microondas excitan las moléculas de agua de la comida, causando que se muevan más y más rápido. ¿Termina ella aquí la explicación, tal como cualquier profesor de ciencias de bachillerato sabe que debería terminar, observando quizás que «la temperatura aumentada *es* justamente energía cinética aumentada de las moléculas constituyentes»? Ella no lo hace. Sigue explicando que porque las moléculas se mueven más rápido, chocan entre sí más a menudo, lo que aumenta la fricción entre las moléculas, y, como todos sabemos, la fricción causa calor. *Betty Crocker todavía piensa que el calor es algo distinto de la energía cinética molecular; algo causado por pero realmente independiente del movimiento molecular*<sup>19</sup>. ¿Por qué los científicos no piensan así también?

Dicho rápidamente, porque las explicaciones de los fenómenos de calor - producción por combustión, por el sol, y en reacciones químicas; de conductividad, incluyendo conductividad en el vacío, la variación de conductividad en materiales distintos, etc. - son *ampliamente* más simples y más coherentes a partir del supuesto de que el calor *es* energía molecular de las moléculas constituyentes. En contraste, intentar hacer que los datos encajen con el supuesto de que el calor es alguna otra cosa *causada por* movimiento molecular acelerado es como intentar clavar gelatina en la pared.

Si alguien está obligado y determinado a adherirse a una termodinámica del calórico, podría, con esfuerzo heroico, lograrlo para sí mismo, aunque los conversos son improbables. El coste, sin embargo, en coherencia con el resto de la teoría científica, por no mencionar otras observaciones, es extremadamente alto. ¿Qué motivaría pagar ese coste? Quizás alguien con voluntad de hierro y juramentado resuelva mantener inmaculada la intuición de que el calor *es lo que es y no otra cosa*. Re-

19 Paul Churchland hizo este descubrimiento en nuestra cocina hace unos ocho años. Nos pareció un caso guay de alguien que realmente no comprende la explicación científica. En vez de pensar a través de la teoría termodinámica, Betty Crocker de manera torpe injerta el caso en una vieja concepción como si tal vieja concepción no necesitase modificación. Alguien que pensase que la electricidad está *causada* por electrones en movimiento contaría un relato comparable al de Betty Crocker: «el voltaje fuerza a los electrones a moverse a través del alambre, y conforme lo hacen causan el aumento de la electricidad estática, y una chispa entonces salta de electrón en electrón en el alambre».

trospectivamente, y sabiendo lo que sabemos ahora, la idea de que alguien iría muy lejos para defender esa intuición del calor parece más bien una pérdida de tiempo.

En el caso que nos ocupa, predigo que la potencia explicativa, la coherencia y la economía favorecerán la hipótesis de que la conciencia *es* justamente algún patrón de actividad de las neuronas. Puedo equivocarme. Si me equivoco, no será porque una intuición basada introspectivamente es inmutable, sino porque la ciencia nos conduce en una dirección diferente. Si tengo razón, y ciertos patrones de actividad cerebral *son* la realidad detrás de la experiencia, este hecho por sí mismo no cambia mi experiencia y de repente me permite (a mi cerebro) considerar mi cerebro como un escáner RM o un neurocirujano podría considerarlo. Continuaré teniendo experiencias en la antigua manera regular, aunque para comprender su realidad neuronal, mi cerebro necesita *tener* montones de experiencias y experimentar montones de aprendizajes.

Finalmente, impidiendo saltar al caballo del dualista, la idea de que tiene que haber un fundamento firme de apariencia subjetiva sobre el cual los descubrimientos de realidad/apariencia tienen que apoyarse es vagamente extraña. Parece un poco como insistir que «abajo» no puede ser relativo a donde uno está en el espacio; abajo es abajo. O como insistir que el tiempo no puede ser relativo, que cualesquiera dos eventos suceden al mismo tiempo o no, y eso es eso. Los humanos son productos de evolución; los sistemas nerviosos han evolucionado en el contexto de la competencia por la supervivencia - en la lucha para tener éxito en las cuatro tareas de alimentarse, huir, luchar y reproducirse. El modelo cerebral del mundo externo mejora mediante la apreciación de variadas distinciones de realidad/apariencia - en pocas palabras, mediante razón crítica común y mediante ciencia. En la naturaleza de las cosas, es muy probable que el modelo cerebral de su mundo interno también permita descubrimientos apariencia/realidad. El cerebro no evolucionó para conocer la naturaleza del sol tal como es conocido por un físico, ni para conocerse a sí mismo tal como es conocido por un neurofisiólogo. Pero, en las circunstancias correctas, puede llegar a conocerlos de algún modo<sup>20</sup>.

20 Véase P. M. Churchland (1993).

#### D. Porque la conciencia es una máquina virtual

Este es el punto de vista de D. C. Dennett<sup>21</sup>. Como Searle, Dennett es no dualista. A diferencia de Searle, quien piensa que mucho, si no todo, acerca de la conciencia puede ser descubierto por la neurociencia, Dennett está convencido desde hace tiempo de que el estudio del cerebro mismo - su fisiología y anatomía - es con mucho una pérdida de tiempo en lo que se refiere a la comprensión de la naturaleza de la conciencia y de la cognición. Simplificado, lo esencial de su idea es esto: los humanos se hacen conscientes conforme adquieren lenguaje y aprenden a hablarse a sí mismos. Lo que sucede en esta transformación es que una máquina paralela (las redes neuronales del cerebro) simula una máquina serial (las operaciones se ejecutan una a una, en una secuencia, según reglas, que pueden ser recursivas).

Mediante la adquisición de un lenguaje y aprendiendo luego a hablar en silencio consigo mismo, uno supuestamente crea una máquina virtual de conciencia en el cerebro. Dennett explica lo que es esto por medio de una analogía fundamental: es como crear una máquina virtual para simular pilotar un aeroplano en tu computador de escritorio instalando software, tal como *Flight Simulator*. La conciencia tiene la misma relación con el cerebro que la simulación de vuelo tiene con los eventos dentro del computador.

La moraleja metodológica de Dennett es inequívoca: al igual que no podemos esperar aprender mucho acerca del simulador de vuelo (su alcance y límites, cómo funciona) estudiando las interioridades del computador mientras hace funcionar *Flight Simulator* tampoco podemos esperar aprender mucho acerca de la conciencia estudiando las interioridades del cerebro mientras es consciente. Si uno quiere saber acerca de *Flight Simulator* y sus muchas propiedades, lo mejor que puede hacer es estudiar su ejecución - en cierto sentido, no hay realmente otra cosa en *Flight Simulator* que su ejecución. Encontramos fructífero al hablar de *Flight Simulator* decir cosas como «su altímetro registra altitud», pero esto no quiere decir que haya algo en mi computador que realmente está alto en el cielo o algo que mida qué alto es realmente. Tal modo de hablar es simplemente una manera económica, conveniente de dar sentido a la ejecución de pantalla del computador cuando hace funcionar el software *Flight Simulator*.

21 D. C. Dennett, *Consciousness Explained* (1991).

Lo mismo (más o menos) en cuanto a la conciencia. El cerebro es el hardware sobre el cual el software de conciencia funciona, con lo que mirar al cerebro mismo no va a enseñarnos mucho acerca del software mismo. Incluso al igual que es erróneo suponer que el computador tiene una pequeña pista oculta plegada dentro que se extiende cuando aprieto un botón, también es erróneo pensar que el cerebro realmente hace algo como rellenar la zona ciega o rellenar mientras vemos movimiento subjetivo (como en una película)<sup>22</sup>. Dennett cree que nos ha mostrado que no hay realmente mucho a explicar en el modo de experiencia interna después de todo. Como con *Flight Simulator*, si quieres saber acerca de la conciencia y sus propiedades, es la ejecución bajo una variedad de condiciones la que necesita ser estudiada. Basado en la ejecución puedes por supuesto inferir las diversas propiedades computacionales del software. *Y eso es todo lo que habrá para explicar la conciencia*. Por tanto, los instrumentos de la psicología experimental bastarán. Los detalles de la neurociencia podrían decirnos algo acerca de cómo funciona el software en el cerebro; eso no nos dirá nada acerca de la naturaleza de la conciencia, sino sólo acerca de cómo el cerebro hace funcionar software. Esto, en resumidas cuentas, es mi comprensión de la convicción que inspira a Dennett en el título de su libro *Consciousness Explained*.

¿En qué medida es plausible el relato de Dennett? Mi crítica aquí se apoya en la obra de Paul Churchland<sup>23</sup> y se centrará principalmente en esta cuestión: ¿es remotamente razonable que cuando somos conciencia la máquina paralela (el cerebro) esté simulando una máquina serial? Como preliminar de archivo, sin embargo, obsérvese que el conjunto de Dennett ha estado sujeto a análisis intenso y cuidadoso. En primer lugar, su pretensión de que la adquisición del lenguaje humano es una condición necesaria de la conciencia humana ha sido repetidamente desafiada y enteramente criticada<sup>24</sup>. Constantemente se ha observado que esto parece implicar que los infantes preverbales no son conscientes; que otros animales, tales como chimpancés y orangutanes, no son conscientes; que sujetos con afasia global o hemisferiotomía izquierda no son conscientes. En breve, la respuesta de Dennett es ésta: así es, los sujetos no verbales no son conscientes en la manera en que un humano plenamente verbal es consciente; por ejemplo, no pueden pensar acerca de si los intereses bajarán el próximo mes. Desafortunadamente, la respuesta de

22 Para una crítica de Dennett sobre el rellenado, véase P. S. Churchland y Ramachandran (1993).

23 Paul Churchland, *The Engine of Reason, the Seat of the Soul* (1995).

24 Véase especialmente Owen Flanagan (1992) y Ned Block (1993).

Dennett es tangencial a la crítica. El asunto es si los niños preverbales y los animales pueden ser conscientes de colores, sonidos, olores, extensión espacial, movimiento, estar mareado, sentir dolor, etc. del mismo modo en que yo soy consciente de ellos.

En segundo lugar, al conceder Dennett un estatus preeminente a la actividad lingüística y «desbancar» correlativamente las experiencias sensoriales (por ejemplo, el rellenado), sentimientos y en general la cognición no lingüística, ha estado sometido a un constante aluvión de quejas<sup>25</sup>. Lamentablemente, sólo puedo dar aquí una versión altamente truncada de los largos y a veces retorcidos debates entre Dennett y varios críticos. El centro de las quejas es que Dennett erróneamente asume que la ejecución es todo lo que necesita explicar - que explicar *informes* de experiencia consciente es equivalente a explicar la experiencia consciente misma. La respuesta básica de Dennett aquí ha sido despachar a los que se quejan como habiendo fracasado en comprenderle apropiadamente, regañándoles por estar todavía en las garras de malos y viejos hábitos conceptuales que implican homúnculos, fantasmas en la máquina, cartesianismo furtivo, y errores relacionados. Baste decir que la postura de Dennett «si no estás de acuerdo has entendido mal», mientras concebiblemente verdadera de algunos críticos, no parece verdadera de todos.

¿Es una «máquina serial virtual» *necesaria* para obtener una ordenación temporal de uno-tras-otro? En absoluto. Por ejemplo, ha sido bien conocido durante al menos ocho años que las redes neuronales con bucles recurrentes pueden proporcionar secuenciación temporal, y lo hacen muy económica y elegantemente<sup>26</sup>. Para un ejemplo reciente, un bello trabajo usando «algoritmos genéticos con valores reales para la evolución de redes neuronales recurrentes de tiempo continuo capaces de conducta y aprendizaje secuencial» ha sido hecho por Randall Beer y otro trabajo de secuenciación ha sido hecho por Michael Mozer<sup>27</sup>. Claramente, tareas de secuenciación *per se* no implican la existencia de una máquina serial simulada<sup>28</sup>.

¿Es una máquina serial virtual necesaria para obtener conducta seguidora de reglas tal como vemos en la ejecución lingüística? En absoluto. De nuevo, como Elman y sus colegas han mostrado, redes neuronales recurrentes pueden conseguir esto muy bien<sup>29</sup>. ¿Se necesita una máqui-

25 Ibid. También John Searle, *The Rediscovery of the Mind* (1992).

26 Véase por ejemplo, Singh (1992), Mozer (1992), Sutton, Mamelak y Hobson (1992).

27 Beer y Gallagher (1992), Beer (1995a), Beer (1995b).

28 Véase nuestra discusión de redes variadas en Churchland y Sejnowski (1992).

29 Elman (1991). Véase también Mozer y Bachrach (1991), Pollack (1991), Giles et al. (1992), Jain (1992), Pinkas (1992), Sumida y Dyer (1992).



na serial virtual para restringir una cierta clase de operaciones a una a la vez? En absoluto. Primero, una clase especial de operaciones podría ser el output de una red, aunque una red ampliamente distribuída. Segundo, podrían ser el output de una interacción ganador-lleva todo entre redes<sup>30</sup>. Y hay montones de otras arquitecturas para realizar esto. El sistema motor probablemente funciona así, pero no hay razón para pensar que *simula* una máquina serial<sup>31</sup>.

En tercer lugar, ¿deberíamos asumir que la conciencia envuelve solamente una operación a la vez? Casi ciertamente no. Concediendo que la capacidad atencional es mucho más pequeña que la capacidad extra-atenta para representar<sup>32</sup>, ¿por qué concluir que podemos atender a sólo una cosa a la vez? Cuando miro un tazón colorado, ¿puedo ver más de un color a la vez? Probablemente.

En cuarto lugar, ¿es la simulación de máquina serial necesaria para permitir propiedades recursivas, tales que uno pueda ser autoconsciente (pensar acerca de lo que justamente uno se dijo a sí mismo)? En absoluto. Las redes neuronales recurrentes son bastante poderosas y bastante complejas para conseguir esto muy bien. En verdad, la recurrencia es probablemente un rasgo clave de varios subsistemas de supervisión en el sistema nervioso, incluyendo la termoregulación. ¿Hay alguna razón fundamental para decir «cuando somos conscientes el cerebro tiene que simular una máquina serial»? No veo ninguna<sup>33</sup>. Esto no entraña que Dennett tiene que estar equivocado, sino sólo que no hay razón para pensar que está en la pista correcta.

#### E. El problema está más allá de nuestra débil inteligencia

Inicialmente, esta pretensión parece ser un reconocimiento modesto de nuestras limitaciones (Colin McGinn 1990). De hecho, es una predicción poderosa basada no sobre pruebas sólidas, sino sobre ignorancia

30 Lange (1992).

31 E. g. Viola, Lisberger y Sejnowski (1992), Berthier, Singh, Barto y Houk (1992).

32 Verghese y Pelli (1992) concluyen a partir de sus estudios que la capacidad del mecanismo atencional se limita a entre 44 y 15 bits por vista momentánea. Calcularon que la capacidad pre-atenta es mucho mayor - unos 2106 bits. Sus datos sobre la capacidad atencional son consistentes con el hecho de prestar atención a (o ser consciente de) más de una cosa a la vez.

33 Rick Grush ha hecho una brillante exploración de la idea de que los cerebros emulan aspectos del mundo, incluyendo el cuerpo, en su tesis doctoral *Emulation and Cognition*. Para una aproximación al problema usando redes recurrentes, véase también Dan Lloyd, «Consciousness: A Connectionist Manifesto» (no publicado).

profunda. Pues por todo lo que podemos estar seguros ahora, la predicción podría ser correcta, pero igualmente podría muy bien ser falsa. ¿En qué medida es débil nuestra inteligencia? ¿En qué medida es difícil el problema? ¿Cómo podrías posiblemente saber que resolver el problema está más allá de nuestro alcance, sin importar cómo se desarrollen la ciencia y la tecnología? Puesto que no es conocido que el cerebro sea más complicado que es inteligente, abandonar la tentativa de averiguar cómo funciona sería decepcionante. Al contrario, mientras los experimentos continúen produciendo resultados que contribuyen a nuestra comprensión, ¿por qué no seguir adelante?<sup>34</sup>

### III. LOCALIZANDO LOS MECANISMOS NEURONALES DE LA CONCIENCIA

#### A. Encontrando una ruta de entrada

En neurociencia existen muchos datos en niveles superiores relevantes para la conciencia. Visión ciega, semidesatención, cerebros divididos, anosognosia (inconsciencia de déficit), para los que comienzan, son exigencias poderosas para guiar la reflexión teórica. Estudios cuidadosos que usan mecanismos de visualización tales como imágenes de resonancia magnética (IRM) y tomografía por emisión de positrones (TEP) nos han permitido relacionar tipos específicos de pérdidas funcionales con regiones cerebrales particulares<sup>35</sup>. Esto ayuda a estrechar el campo de las estructuras que podemos seleccionar para una preliminar microexploración<sup>36</sup>.

Por ejemplo, el hipocampo podría haber parecido un candidato probable para un papel central en la conciencia porque es una región de tremenda convergencia de fibras desde diversas áreas del cerebro. Sabemos ahora, sin embargo, que la pérdida bilateral del hipocampo, aunque daña la capacidad para aprender nuevas cosas, no entraña pérdida de conciencia. En este estadio, excluir algo es por sí mismo un avance valioso. Sabemos también que ciertas estructuras del tallo cerebral tales como el locus ceruleus (LC) son indirectamente necesarias, pero no son

<sup>34</sup> Véase la convincente y más detallada discusión del rechazo de McGinn por Owen Flanagan, Flanagan (1992).

<sup>35</sup> Véase especialmente H. Damasio y A. R. Damasio (1990), H. Damasio (1991), A. R. Damasio (1994), Farah (1994).

<sup>36</sup> Una posibilidad explorada por Penrose (1994) es que la conciencia es un fenómeno mecánico-cuántico, producido en estructuras orgánicas muy pequeñas llamadas microtúbulos. Para una crítica de tal hipótesis, véase Grush y Churchland (1995).

parte del mecanismo de la conciencia. LC juega efectivamente un papel no específico en la vigilia, pero no un papel específico en la conciencia de contenidos particulares, tal como conciencia en un momento dado del color del cielo de la mañana en vez del sonido de los aspersores del césped. Los datos pueden ser fascinantes por sí mismos, pero la cuestión permanece: ¿cómo podemos llegar desde una serie impresionante de datos intrigantes a explicaciones genuinas de los mecanismos básicos? ¿Cómo podemos *empezar*?

Al pensar sobre este problema, he estado muy influenciada por Francis Crick. Su aproximación básica es sencilla: si vamos a resolver el problema, deberíamos tratarlo como un problema científico a ser abordado de la misma manera que abordamos otros problemas científicos difíciles. Como con cualquier misterio científico, lo que queremos es una entrada experimental reveladora. Queremos encontrar un hilo que, al tirar de él, soltará todo un ovillo. Para realizar esto, necesitamos diseñar hipótesis contrastables que puedan conectar macroefectos con microdinámica.

En términos escuetos, a lo que nos enfrentamos es a un problema de satisfacción de exigencia: encontrar fenómenos psicológicos que (a) han sido razonablemente bien estudiados por la psicología experimental, (b) están limitados por datos de lesiones de pacientes humanos al igual que datos de microlesiones precisas de animales, (c) son conocidos como relacionados con regiones cerebrales donde se ha hecho buena neuroanatomía y neurofisiología, y (d) donde sabemos mucho acerca de conectividad con otras regiones cerebrales. La hipótesis de trabajo es que si una persona es consciente de un estímulo, su cerebro será diferente de un modo descubrible respecto del estado en que está despierto y atento pero inconsciente del estímulo. Una estrategia propicia es dar con estas diferencias, guiados por datos de estudios de lesiones, exámenes TEP, estudios de magnetoencefalografía (MEG), etc. El descubrimiento de esas diferencias, en el contexto en general de los datos neurobiológicos, ayudaría a descubrir una teoría del mecanismo.

La idea central es generar una teoría forzada por datos en muchos niveles de organización cerebral - suficientemente forzada de manera que pueda ser sometida a pruebas significativas. En última instancia una teoría de la conciencia necesitará abarcar un campo de procesos envueltos en la conciencia, incluyendo atención y memoria a corto plazo. Inicialmente, sin embargo, puede elegir como objetivo un subconjunto, tal como integración a través del espacio y a través del tiempo. Si la teoría cae en prueba refutadora o si sobrevive a pruebas duras, aprenderemos algo. Es decir, o bien habremos excluido posibilidades específicas - un buen premio en los primeros estadios de la comprensión - o bien pode-

Crick es que (1) para conciencia sensorial, tal como conciencia visual, las cortezas primitivas son fundamentales (por ejemplo, las áreas visuales V2, V3, V5). Esto da sentido a los datos de lesiones, al igual que a los datos TEP recientes (Damasio, Grabowski et al. 1993, Kosslyn et al. 1993) y a los datos de células aisladas (Logothetis y Schall 1989). (2) Dentro de las áreas corticales sensoriales primitivas, las células piramidales en la capa 5 y posiblemente en la capa 6 juegan el papel central.

¿En qué medida es buena esta idea? Parte de su atractivo es su pie firme en estructura básica. En biología, la solución a problemas difíciles acerca de mecanismos puede ser enormemente facilitada mediante la identificación de estructuras críticas. Crudamente, si sabes «qué», ello ayuda enormemente a imaginar «cómo». Por sí misma, la hipótesis de Crick sólo puede ser una pequeña pieza del rompecabezas. Si somos afortunados, sin embargo, ella, o algo como ella, puede ser una pieza *central* del rompecabezas. Este no es el momento para una discusión más completa de esta hipótesis. Baste decir que verdadera o falsa, la hipótesis de Crick proporciona una audaz ilustración de cómo aproximarse a un problema espinoso que es a menudo desechado como inaccesible.

## 2. La hipótesis de Llinás

Otra prometedora ruta de entrada es sugerida por las diferencias - fenomenológicas y neurobiológicas - entre los estados dormir/ sueño/ despierto (DSD)<sup>38</sup>. Este punto de entrada es atractivo, en primer lugar, porque se da la familiar y dramática pérdida de conciencia en el dormir profundo, que es recuperada cuando nos despertamos, y está también presente probablemente mientras soñamos. El fenómeno está muy disponible en montones de sujetos diferentes y a través de muchas especies. En segundo lugar, las técnicas MEG y EEG revelan rasgos cerebrales globales característicos de los diferentes estados. Datos de lesiones humanas y animales son importantes, especialmente en cuanto concierne a déficits en conciencia durante la vigilia. Aquí de nuevo hago observar la importancia de la investigación sobre visión ciega, semidesatención (tendencia a ser inconsciente de los estímulos de varias modalidades en el lado izquierdo del cuerpo), simultanagnosia (incapacidad para ver varias cosas simultáneamente), anosognosia (inconsciencia de

38 Véase también mi discusión en P. S. Churchland (1988).

déficits tales como parálisis), negación de ceguera, inconsciencia de afasia de jerga (de no tener sentido), etc.

En tercer lugar, hemos aprendido muchísimo a partir de las anomalías en y con la manipulación del ciclo DSD y la relación con propiedades cerebrales específicas. En cuarto lugar, algunos de los cambios globales de estado en el ciclo DSD vistos mediante macrotécnicas han estado relacionados mediante microtécnicas con interacciones entre circuitos específicos en el córtex y circuitos subcorticales, especialmente circuitos en varias estructuras centrales del tálamo. En quinto lugar, y más específicamente, los datos MEG revelan una fuerte forma de onda de 40 hercios durante la vigilia y al soñar<sup>39</sup>. La definición y amplitud es muy atenuada mientras se duerme, y la amplitud es modulada mientras se está despierto o soñando. El análisis de la onda mediante MEG revela que es una onda viajera, que se mueve desde la dirección anterior a la posterior en el cerebro, cubriendo la distancia en aproximadamente 12 a 13 *milisegundos*. Datos celulares sugieren que estas propiedades dinámicas emergen de circuitos neuronales particulares y de sus propiedades dinámicas.

¿Qué quiere decir todo esto? Basados en estos datos, y atentos a los varios datos de nivel superior, Rodolfo Llinás y sus colegas (1991, 1993) han formulado la hipótesis de que la organización fundamental que favorece la conciencia y los movimientos vistos en el patrón de DSD son pares de osciladores acoplados, cada uno de los cuales conecta el tálamo y el córtex, pero cada uno conecta poblaciones de células distintas mediante su propio estilo de conectividad distintivo (Figura 2). Una 'familia' de osciladores conecta neuronas de una estructura talámica conocida como los núcleos intralaminares, una constelación de estructuras con forma de pan cuyas neuronas alcanzan las capas superiores del córtex para proporcionar un recubrimiento en forma de ventilador altamente regular del manto cortical entero. La otra 'familia' de osciladores conecta neuronas de los núcleos talámicos para información específica de modalidad (núcleos EM) originando, por ejemplo para la retina o para la cóclea, áreas corticales con modalidad especializada (p. ej. V2, S2). Durante el dormir profundo, las neuronas intralaminares que se proyectan al córtex cesan su conducta de 40 hercios. Durante el dormir profundo y al soñar, las señales externas al córtex están cerradas mediante el núcleo reticular del tálamo. Joseph Bogen (1993) también formula la hipótesis de un papel crucial para las estructuras intralaminares, notan-

39 Véase Llinás y Paré (1991).

do especialmente el amplio abanico desde esos núcleos al córtex, y la fuerte conexión con el estriado.

Siempre tan toscamente, la idea es que la segunda 'familia' de osciladores proporciona el contenido (visual, somatosensorial, etc.) mientras que la primera proporciona el contexto integrador. En el dormir profundo los osciladores están desacoplados; al soñar están acoplados pero el circuito oscilante EM es ampliamente no sensible a las señales externas desde la periferia; en la vigilia, los osciladores están acoplados, y el circuito EM es sensible a señales externas.

¿Cuáles son los efectos en humanos de lesiones en las estructuras talámicas intralaminares (en forma de pan)? En primer lugar, es muy improbable que una lesión ocurriera exclusivamente en los núcleos intralaminares, perdonando a otras estructuras talámicas. Teniendo esa cautela en mente, hago observar que el resultado principal de pequeñas lesiones unilaterales que se cree que residen principalmente en los núcleos intralaminares es desatención (inconsciencia) de todos los estímulos que se originan en el lado corporal opuesto (Watson y Heilman 1979, Watson, Valenstein y Heilman 1981). Las lesiones bilaterales aparentemente producen impasibilidad general, indicando más o menos que el paciente no inicia conducta alguna y responde muy pobremente a estímulos sensoriales o preguntas (Castaigne et al. 1981, Guberman y Stuss 1983). Los escasos estudios con animales que existen son consistentes con los datos humanos (Henderson, Alexander y Nalser 1982, Watson, Miller y Heilman 1978). Bogen (1993) también señala que los datos de lesiones humanas, junto con los patrones de conectividad, hacen a los núcleos intralaminares un candidato para «el donde de la conciencia». Estos datos son importantes y proporcionan un punto de partida útil, pero ulteriores estudios, especialmente usando IRM funcional para localizar áreas de actividad relativa baja y alta, son deseables.

Lesiones en regiones del tálamo específicas de modalidad, en contraste, conducen a pérdidas de conciencia específicas de modalidad - la conciencia visual, por ejemplo, se perderá, pero la conciencia de sonidos, tacto, etc. puede ser normal. De modo intrigante, las MEG de pacientes de Alzheimer que han degenerado a un estado de inanición muestran una forma de onda ruinosa de 40 hercios cuando existe. Obviamente estos datos no son decisivos, pero al menos son consistentes con la hipótesis.

¿Encajan la hipótesis de Llinás y la hipótesis de Crick? Al menos son consistentes. Adicionalmente, se apoyan mutuamente a los niveles neuronal y de red. Un punto alentador es éste: las dos familias de osciladores (EM e intralaminar) se conectan abundantemente entre sí

principalmente en la *capa cortical 5* (Figura 2). A partir de lo cual podemos decir ahora que esas conexiones parecen ser los medios principales por los que se acoplan los osciladores. La posibilidad considerada aquí es que la sincronía temporal que Crick supone en las neuronas que llevan señales acerca de los estímulos externos puede ser orquestada mediante el circuito intralaminar-cortical. Las conexiones entre las estructuras del tallo cerebral y el núcleo intralaminar podían tener un papel en modular la vigilia y la alerta.

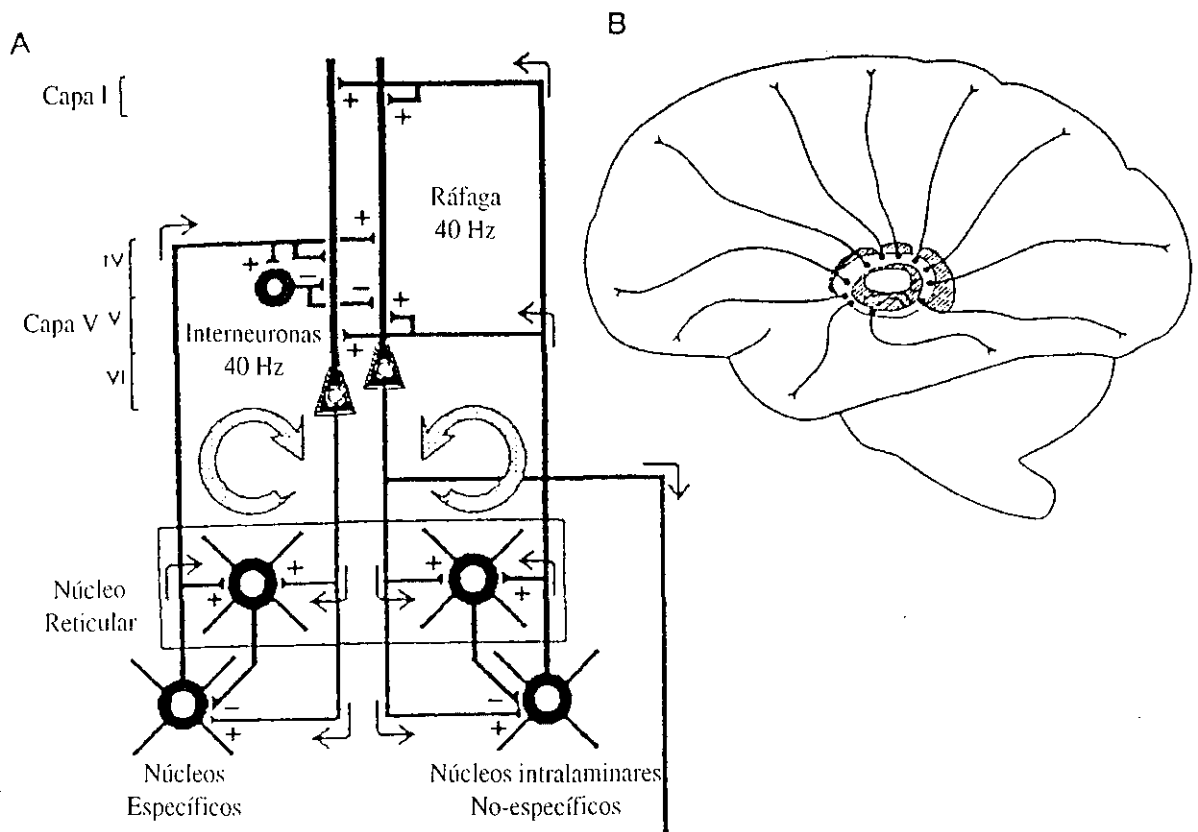


Figura 2

Diagrama esquemático de los circuitos entre el tálamo y el córtex cerebral propuesto para servir al enlace temporal. (A) Diagrama de dos tipos diferentes de circuito que conectan tálamo y córtex. A la izquierda, núcleos sensoriales o núcleos motores específicos del tálamo se proyectan a la capa 4 del córtex, produciendo oscilación cortical mediante activación directa e inhibición de alimentación hacia adelante via interneuronas inhibitorias de 40 hercios. De manera colateral estas proyecciones producen retroalimentaciones talámicas via el núcleo reticular (una especie de piel que cubre el tálamo). La senda de retorno se proyecta de nuevo a los núcleos reticulares y específicos via células de la capa 6. A la derecha, el segundo bucle muestra a los núcleos intralaminares no-específicos del tálamo que se proyectan a la capa 5 del córtex, produciendo una ráfaga de 40 Hz. Este bucle también incluye retroalimentación talámica a través del núcleo reticular.

cleos intralaminares no específicos proyectándose a la capa 1 del córtex, con efectos colaterales en el núcleo reticular. Las células de la capa 5 devuelven la oscilación a los núcleos intralaminar y reticular, estableciendo un segundo bucle de resonancia. La conjunción de los bucles específico y no específico es propuesta para generar el enlace temporal. La conectividad entre los bucles se aprecia principalmente en la capa 5. (B) Diagrama esquemático mostrando el núcleo intralaminar como una masa neuronal circular (zona de puntos). Otras partes del tálamo se muestran en zona sombreada. El núcleo intralaminar se proyecta ampliamente a través del córtex a la capa 1. (Tomado de Llinás y Ribary 1993).

Muchas cuestiones surgen ahora por sí mismas. Por ejemplo, ¿cómo se comunican las estructuras fundamentales de la conciencia con la conducta? Más específicamente, ¿cuáles son las conexiones entre los núcleos intralaminares y las estructuras motoras, y entre la capa 5 de las cortezas sensoriales y las estructuras motoras; tienen las proyecciones desde los núcleos intralaminares al córtex cingulado un papel en la atención? Estas son cuestiones motivadas por datos independientes. La convergencia de hipótesis es por supuesto alentadora, pero está bien recordar que puede también alentarnos hacia el engaño. La sabiduría aconseja optimismo cauteloso.

¿Pero falta algo aquí? Probablemente. Como Kant podría haber dicho a Hume, el cerebro no producirá conciencia a menos que el sistema nervioso también genere una representación del yo - una representación que conlleva lo que llamaríamos «un punto de vista». Y ésta es en verdad precisamente la hipótesis ofrecida por Antonio Damasio (1994). Según la perspectiva de Damasio, los mecanismos neurobiológicos para la conciencia visual, por ejemplo, están esencialmente interconectados con los mecanismos para representarse a sí mismo como una cosa que tiene experiencias, que siente, recuerda y planea; como una cosa que ocupa espacio y que perdura a través del tiempo. Suponer que la conciencia visual puede comprenderse al margen de la autorepresentación es como suponer que la evolución puede comprenderse al margen del ambiente.

Las ideas de Damasio a este respecto han surgido de muchos años de observar pacientes con el cerebro dañado, y de reflexionar sobre las maneras en que la conciencia se relaciona con la autorepresentación y sobre cómo esto a su vez se relaciona con la representación del cuerpo. (Para los detalles de su hipótesis, véase su libro *Descartes' Error*, 1994). Contra un telón de fondo de neuroanatomía y neurofisiología básicas, Damasio ve la complejidad representacional y la interdependencia como elementos clave para explicar la conciencia. Toscamente, la sincronía de disparo en ciertos circuitos visuales puede ser una condición *necesaria*



*ria* para la conciencia visual, pero ciertamente no es una condición *suficiente*. Construir una hipótesis plausible para capturar una condición necesaria es bastante duro; identificar las condiciones neuronales adicionales que conjuntamente son necesarias y suficientes será más duro todavía. Incluso así, la idea central de Damasio es a la vez poderosa y razonable: la *representación del cuerpo*, que sistemáticamente integra estimulación corporal e información del estado del cuerpo, proporciona un andamiaje para la *autorepresentación*, y la autorepresentación es el punto de anclaje de la conciencia - específica de modalidad y de otras maneras. Por otra parte las ventajas adaptativas para la vida de una representación del cuerpo integrada son muy obvias (véase asimismo Damasio 1994). Uno sólo necesita pensar en los machos de mono vervet subdominantes moviéndose furtivamente por la noche para aparearse con hembras mientras el macho principal dormita, para darse cuenta de la importancia de «saber qué camino está libre», para poner el conocimiento del propio yo corporal en su forma más general.

Mientras Kant, un testarudo antireduccionista, estaba convencido de que la naturaleza del yo era para siempre empíricamente no investigable, Damasio encuentra lugares donde el progreso científico es posible; mientras Kant pensaba en el yo en términos de una muy misteriosa «unidad transcendental de la apercepción», Damasio le da una base tranquilizadora concreta en términos de representación neuronal del cuerpo: la piel, músculos, articulaciones, vísceras, etc. Como otras representaciones cerebrales, la naturaleza de la representación del cuerpo es investigable con las técnicas combinadas de la neuropsicología, la neurobiología y la modelación de redes neuronales. Y si Damasio tiene razón, entonces los mecanismos neuronales de la autorepresentación son investigables también.

Veo las tres aproximaciones ampliamente neurocientíficas - las de Crick, Llinás y Damasio - como estrategias complementarias para encarar diferentes pero superpuestos segmentos de un problema amplio y desconcertante. Cada una tiene fuerzas particulares con relación a algún aspecto del problema, y cada una presenta preguntas y desafíos a las otras - el auténtico dechado de descubrimiento co-evolucionista.

#### IV. OBSERVACIONES FINALES

Al considerar los temas desde el lado misterioso de un fenómeno, las soluciones pueden parecer imposibles, y quizás incluso no deseadas. Por el lado de la comprensión, sin embargo, las soluciones parecen casi obvias y difíciles de no encontrar. Uno podría preguntarse ¿porqué llevó

tanto tiempo entender qué son los elementos? ¿Cómo pudo alguien tan brillante como Aristóteles no encontrar la plausibilidad en la idea de Aristarco de que la Tierra era una esfera que se movía alrededor del sol? Las verdades más profundas son todas demasiado fáciles de no encontrar por supuesto, al igual que es demasiado fácil para nosotros no encontrar, sea lo que sea, lo que explica por qué los animales duermen y sueñan, y qué es el autismo. Los problemas para la neurociencia y la psicología experimental son duros, pero conforme vamos avanzando y conforme las nuevas técnicas aumentan el acceso no dañino a los procesos cerebrales globales en los humanos, las intuiciones cambian. Lo que nos parece obvio fué noticia caliente y sorprendente hace sólo una generación; lo que parece confundir nuestra imaginación es de modo rutinario aceptable por la nueva cohorte de estudiantes graduados. ¿Quién puede decir con certeza si o no todas nuestras cuestiones acerca de la conciencia pueden eventualmente contestarse? Mientras tanto, es recompensador ver progreso - ver algunas cuestiones cambiar de situación, desde Misterios que sólo podemos contemplar con temor reverencial a Problemas duros que estamos comenzando a descifrar.

#### REFERENCIAS

- BEER, R. D. (1995a), «A dynamical systems perspective on agent-environment interaction», *Artificial Intelligence* , 72, 173-215.
- BEER, R. D. (1995b), «On the dynamics of small continuous-time recurrent neural networks», *Adaptive Behavior* , 3, 471-511.
- BEER, R. D. y J. C. GALLAGHER (1992), «Evolving dynamical neural networks for adaptive behavior», *Adaptive Behavior* , 1, 91-122.
- BERTHIER, N. E., S. P. SINGH, A. G. BARTO y J. C. HOUK (1992), «A cortico-cerebellar model that learns to generate distributed motor commands to control a kinematic arm», J. E. Moody, S. J. Hanson y R. P. Lippmann (eds.), *Neural Information Processing Systems* , 4ª ed., San Mateo, Ca.: Morgan-Kaufman, 611-618.
- BICKLE, J. (1992), «Revisionary physicalism», *Biology and Philosophy* , 7, 411-430.
- BLOCK, Ned (1993), «Consciousness ignored? Review of D. C. Dennett's *Consciousness Explained*», *Journal of Philosophy* , 90, 4, 83-91.
- BOGEN, J. E. (1993), «Intralaminar nuclei and the where of awareness», *Society for Neuroscience Abstracts* .
- CASTAIGNE, P. F., LHERMITTE, A. BUGÉ, R. ESCOUROLLE, J. J. HAUW y O. LYONCAEN (1981)»Paramedian thalamic and midbrain infarcts: Clinical and neuropathological study», *Annals of Neurology* , 10, 127-148.

- CHURCHLAND, Paul M. (1981), «Eliminative Materialism and the Propositional Attitudes», *Journal of Philosophy*, 78, 2, 67-90.
- CHURCHLAND, P. M. (1988), *Matter and Consciousness*, 2ª ed., Cambridge, Mass.: MIT Press.
- CHURCHLAND, P. M. (1989), *A Neurocomputational Perspective. The Nature of Mind and the Structure of Science*, Cambridge, Mass.: MIT Press.
- CHURCHLAND, P. M. (1993), «Evaluating our self conception», *Mind and Language*, 8, 211-222.
- CHURCHLAND, P. M. (1994), «Betty Crocker's theory of the mind: A review of *The rediscovery of the Mind* by John Searle», *London Review of Books*, mayo, 1994.
- CHURCHLAND, P. M. y P. S. CHURCHLAND (1990), «Intertheoretic reduction: A neuroscientist's field guide», *Seminars in the Neurosciences*, 4, 249-256.
- CHURCHLAND, P. S. (1986), *Neurophilosophy*, Cambridge, Mass.: MIT Press.
- CHURCHLAND, P. S. (1987), «Replies to Comments. Symposium on Patricia Smith Churchland's *Neurophilosophy*», *Inquiry*, 29, 241-272.
- CHURCHLAND, P. S. (1988), «Reduction and the neurobiological basis of consciousness», A. J. Marcel y E. Bisiach (eds.), *Consciousness in Contemporary Science*, New York: Oxford University Press, 273-304.
- CHURCHLAND, P. S. (1994), «Can neurobiology teach us anything about consciousness?», Presidential Address to the American Philosophical Association, Pacific Division, *Proceedings and Addresses of the American Philosophical Association*, Lancaster, Pa.: Lancaster Press, 67-4, 23-40.
- CHURCHLAND, P. S. y T. J. SEJNOWSKI (1989), «Brain and cognition», M. Posner (ed.), *Foundations of Cognitive Science*, Cambridge, Mass.: MIT Press, 245-300.
- CHURCHLAND, P. S. y T. J. SEJNOWSKI (1992), *The Computational Brain*, Cambridge, Mass.: MIT Press.
- CHURCHLAND, P. S. y V. S. RAMACHANDRAN (1993), «Filling-in: Why Dennett is wrong», B. Dahlbom (ed.), *Dennett and his Critics*, Oxford: Blackwell, 28-52.
- CRICK, F. H. C. (1994), *The Astonishing Hypothesis*, New York: Scribner's Sons.
- CRICK, F. H. C. y C. KOCH (1990), «Towards a neurobiological theory of consciousness», *Seminars in the Neurosciences*, 4, 263-276.
- DAMASIO, A. R. (1994), *Descartes' Error*, New York: Putnam.
- DAMASIO, H. (1991), «Neuroanatomy of frontal lobe in vivo: A comment on methodology», H. Levin, H. Eisenberg y A. Benton (eds.), *Frontal Lobe Function and Dysfunction*, New York: Oxford University Press, 92-121.

- DAMASIO, H. y A. R. DAMASIO (1990), «The neural basis of memory, language and behavioral guidance: Advances with the lesion method in humans», *Seminars in the Neurosciences* , 4, 277-286.
- DAMASIO, H., T. J. GRABOWSKI et al. (1993), «Visual recall with eyes closed and covered activated early visual cortices», *Society for Neuroscience Abstracts* (658.4)
- DENNETT, D. C. (1991), *Consciousness Explained* , Boston: Little, Brown and Co.
- ELMAN, J. L. (1991), «Distributed representations, simple recurrent networks, and grammatical structure», *Machine Learning* , 7, 195-225.
- FARAH, M. J. (1994), «Neuropsychological inference with an interactive brain: A critique of the 'locality assumption'», *Behavioral and Brain Sciences* , 17, 43-104.
- FEYERABEND, P. K. (1981), *Philosophical Papers* , vol. 1 y 2, Cambridge: Cambridge University Press.
- FODOR, J. (1990), *A Theory of Content and Other Essays* , Cambridge, Mass.: MIT Press.
- FLANAGAN, O. (1992), *Consciousness Reconsidered* , Cambridge, Mass.: MIT Press.
- FLANAGAN, O. (1996), «Prospects for a unified theory of consciousness, or, what dreams are made of», J. Cohen y J. Schooler (eds.), *Scientific Approaches to the Question of Consciousness: 25th Carnegie Symposium on Cognition* , Hillsdale, N. J.: L. Erlbaum.
- GRUSH, Rick (no publicado), *Emulation and Cognition* , tesis de Ph. D. en la UCSD.
- GRUSH, Rick y P. S. CHURCHLAND (1995), «Gaps in Penrose's toiling», *Journal of Consciousness Studies* , 2, 10-29.
- GILES, C. L., C. B. MILLER, D. CHEN, G. Z. SUN, H. H. CHEN y Y. C. LEE (1992), «Extracting and learning and unknown grammar with recurrent neural nets», J. E. Moody, S. J. Hanson y R. P. Lippmann (eds.), *Neural Information Processing Systems 4* , San Mateo, Ca.: Morgan-Kaufman, 317-324.
- GUBERMAN, A. y D. STUSS (1993), «The syndrome of bilateral paramedian thalamic infarction», *Neurology* , 33, 540-546.
- HENDERSON, V. W., M. P. ALEXANDER y M. A. NALSER (1982), «Right thalamic injury, impaired visuospatial perception and alexia», *Neurology* , 32, 235-240.
- JAIN, A. N. (1992), «Generalizing performance in PARSEC. A structured connectionist parsing architecture», J. E. Moody, S. J. Hanson y R. P. Lippmann (eds.), *Neural Information Processing Systems 4* , San Mateo, Ca.: Morgan-Kaufman, 209-216.

- KOSSLYN, S. M., N. M. ALPERT, W. L. THOMPSON, V. MALJKOVIC, S. B. WEISE, C. F. CHABRIS, S. E. HAMILTON, S. L. RAUCH y F. S. BUOANNO (1993), «Visual mental imagery activated topographically organized visual cortex: PET investigations», *Journal of Cognitive Neuroscience* , 5, 263-287.
- LANGE, T. E. (1992), «Dynamically-adaptive winner-take-all networks», J. E. Moody, S. J. Hanson y R. P. Lippmann (eds.), *Neural Information Processing Systems 4* , San Mateo, Ca.: Morgan-Kaufman, 341-348.
- LLINAS, R. R. y D. PARE (1991), «Of dreaming and wakefulness», *Neuroscience* , 44, 521-535.
- LLINAS, R. R. y U. RIBARY (1993), «Coherent 40-Hz oscillation characterizes dream state in humans», *Proceedings of the National Academy of Sciences*, 90, 2078-2081.
- LOGOTHETIS, N. y J. D. SCHALL (1989), «Neural correlates of subjective visual perception», *Science* , 245, 753-761.
- LYCAN, W. G. (1987), *Consciousness* , Cambridge, Mass.: MIT Press.
- MCGINN, C. (1990), *The Problem of Consciousness* , Oxford: Blackwell.
- MOZER, M. C. (1992), «Induction of multiscale temporal structure», J. E. Moody, S. J. Hanson y R. P. Lippmann (eds.), *Neural Information Processing Systems 4* , San Mateo, Ca.: Morgan-Kaufman, 275-282.
- MOZER, M. C. y J. BACHRACH (1991), «SLUG: A connectionist architecture for inferring the structure of fine-state environments», *Machine Learning*, 7, 139-160.
- PENROSE, Roger (1994), *Shadows of the Mind* , Oxford: Oxford University Press.
- PINKAS, G. (1992), «Constructing proofs in symmetric networks», J. E. Moody, S. J. Hanson y R. P. Lippmann (eds.), *Neural Information Processing Systems 4* , San Mateo, Ca.: Morgan-Kaufman, 217-224.
- POLLACK, J. B. (1991), «The induction of dynamical recognizers», *Machine Learning* , 7, 227-252.
- SCHAFFNER, K. F. (1993), «Theory structure, reduction, and disciplinary integration in biology», *Biology and Philosophy* , 8, 319-348.
- SEARLE, J. R. (1992), *The Rediscovery of the Mind* , Cambridge, Mass.: MIT Press.
- SINGH, S. P. (1992), «The efficient learning on multiple task sequences», J. E. Moody, S. J. Hanson y R. P. Lippmann (eds.), *Neural Information Processing Systems 4* , San Mateo, Ca.: Morgan-Kaufman, 251-258.
- SUMIDA, R. A. y M. G. DYER (1992), «Propagation filters in PDS networks for sequencing and ambiguity resolution», J. E. Moody, S. J. Hanson y R. P. Lippmann (eds.), *Neural Information Processing Systems 4*, San Mateo, Ca.: Morgan-Kaufman, 233-240.

- SUTTON, J. P., A. N. MAMELAK y J. A. HOBSON (1992), «Network model of state-dependent sequencing», J. E. Moody, S. J. Hanson y R. P. Lippmann (eds.), *Neural Information Processing Systems 4*, San Mateo, Ca.: Morgan-Kaufman, 283-290.
- VERGHESE, P. y D. G. PELLI (1992), «The information capacity of visual attention», *Vision Research*, 32, 983-995.
- VIOLA, P. A., S. G. LISBERGER y T. J. SEJNOWSKI (1992), «Recurrent eye tracking network using a distributed representation of image motion». J. E. Moody, S. J. Hanson y R. P. Lippmann (eds.), *Neural Information Processing Systems 4*, San Mateo, Ca.: Morgan-Kaufman, 380-387.
- WATSON, R. T. y K. M. HEILMAN (1979), «Thalamic neglect», *Neurology*, 29, 690-694.
- WATSON, R. T., B. D. MILLER y K. M. HEILMAN (1978), «Nonsensory neglect», *Annals of Neurology*, 3, 505-508.
- WATSON, R. T., E. VALENSTEIN y K. M. HEILMAN (1981), «Thalamic neglect: Possible role of the medial thalamus and nucleus reticularis in behavior», *Archives of Neurology*, 38, 501-506.