

Traducción de Carmen Martínez Gimeno

# Índice

Prólogo

1. Inteligencia Artificial

2. Redes Neuronales

3. El cerebro Humano

4. La Memoria

5. Un Nuevo Marco para la Inteligencia

6. Cómo Funciona la Corteza Cerebral

7. Conciencia y Creatividad

8. El Futuro de la Inteligencia

Epílogo

Apéndice: Predicciones Verificables

Bibliografía

Agradecimientos

# **Sobre la Inteligencia**

## Prólogo

**Dos** pasiones animan este libro y mi vida.

Durante veinticinco años he sido un entusiasta de los ordenadores portátiles. En el mundo de alta tecnología de Silicon Valley se me conoce por haber puesto en marcha dos empresas, Palm Computing y Handspring, y por ser el creador de muchos ordenadores de mano y teléfonos móviles como la PalmPilot y el Treo.

Pero tengo una segunda pasión que precede a mi interés por los ordenadores y que considero más importante: los cerebros me vuelven loco. Quiero entender cómo funcionan no solo desde una perspectiva filosófica, sino al modo detallado de la ingeniería, analizando sus elementos básicos. Mi deseo no se limita a comprender qué es la inteligencia y cómo actúa el cerebro, sino que pretendo aprender a construir máquinas que funcionen de la misma manera. Quiero construir máquinas verdaderamente inteligentes.

La cuestión de la inteligencia es la última gran frontera terrestre de la ciencia. La mayoría de los temas científicos importantes tratan de lo muy pequeño, lo muy grande o de acontecimientos que ocurrieron hace miles de millones de años. Pero toda persona tiene un cerebro. Eres tu cerebro. Si quieres comprender por qué te sientes como te sientes, cómo percibes el mundo, por qué cometes errores, cómo eres capaz de ser creativo, por qué la música y el arte sirven de inspiración, en definitiva, qué es ser humano, necesitas entender el cerebro. Además, una teoría satisfactoria sobre la inteligencia y la función cerebral proporcionará amplios beneficios sociales, que no se limitarán a ayudarnos a curar las enfermedades relacionadas con el cerebro. Seremos capaces de construir máquinas realmente inteligentes, aunque no se parecerán en absoluto a los robots de la ficción popular y la fantasía de la ciencia informática. En su lugar, las máquinas inteligentes surgirán de un nuevo conjunto de principios sobre la naturaleza de la inteligencia. Dichas máquinas nos ayudarán a acelerar nuestro conocimiento del entorno, nos asistirán en la exploración del Universo y lograrán que el mundo sea más seguro. Y de paso se creará una industria enorme.

Por suerte, vivimos en una época en la que es posible resolver el problema que supone la comprensión de la inteligencia. Nuestra generación tiene acceso a una montaña de datos sobre el cerebro reunidos durante cientos de años, y el ritmo al que seguimos reuniéndolos se está acelerando. Solo en Estados Unidos hay miles de neurocientíficos. Sin embargo, no contamos con teorías productivas sobre qué es la inteligencia o cómo funciona el cerebro en su conjunto. La mayoría de los neurobiólogos no piensan mucho sobre teorías generales del cerebro porque están enfrascados en realizar experimentos para reunir más datos sobre los muchos subsistemas cerebrales. Y aunque legiones de programadores informáticos han intentado hacer inteligentes los ordenadores, han fracasado. Y creo que seguirán fracasando mientras sigan ignorando las diferencias entre ordenadores y cerebros.

Así pues, ¿en qué consiste la inteligencia que posee el cerebro, pero no los ordenadores? ¿Por qué un niño de seis años puede saltar con desenvoltura de roca en roca en un arroyo, mientras que los robots más avanzados de nuestra era son zombis pesados? ¿Por qué los niños de tres años van camino de dominar el lenguaje, mientras que los ordenadores resultan incapaces de hacerlo a pesar de medio siglo de esfuerzos de los programadores? ¿Por qué podemos decir que es un gato y no un perro en una fracción de segundo, mientras que un superordenador no es capaz de establecer la distinción? Son grandes misterios que esperan una respuesta. Tenemos abundantes pistas; ahora lo que necesitamos son unas cuantas percepciones decisivas.

Tal vez se pregunten por qué un diseñador de ordenadores escribe un libro sobre cerebros. O expresado de otro modo: si me gustan los cerebros, por qué no desarrollé mi carrera en la ciencia cerebral o en inteligencia artificial. La respuesta es que lo intenté varias veces, pero me negué a estudiar el problema de la inteligencia como lo habían hecho otros antes que yo. Creía que el mejor modo de resolver este problema es emplear la biología detallada del cerebro como una limitación y una pauta, pero concibiendo la inteligencia como un problema computacional, una postura que se sitúa en un punto entre la biología y la ciencia informática. Muchos biólogos tienden a rechazar o ignorar la idea de pensar en el cerebro en términos computacionales, y los científicos informáticos no suelen creer que tengan nada que aprender de la biología. Además, el mundo de la ciencia acepta menos el riesgo que el empresarial. En las empresas tecnológicas, una persona que persigue una nueva idea con un planteamiento razonable puede ascender en su carrera sea cual fuere el éxito que alcance dicha idea. Muchos empresarios lograron triunfar después de varios fracasos previos. Pero en el mundo académico, un par de años dedicados a una nueva idea que no da buenos resultados pueden arruinar para siempre una carrera que se inicia. Por lo tanto, me dediqué a las dos pasiones de mi vida de forma simultánea, creyendo que el éxito en la industria me ayudaría a dedicarme a la ciencia que deseaba; además, necesitaba aprender cómo afecta el cambio en el mundo y cómo vender nuevas ideas, todo lo cual esperaba obtener trabajando en Silicon Valley.

En agosto de 2002 puse en marcha un centro de investigación, el Redwood Neuroscience Institute (RNI), dedicado a la teoría sobre el cerebro. Existen muchos centros de neurociencia en el mundo, pero ningún otro se dedica a la comprensión teórica general de la corteza cerebral, la parte del cerebro humano responsable de la inteligencia. Eso es lo único que estudiamos en el RNI, que en muchos aspectos se asemeja a una empresa que empieza. Perseguimos un sueño que algunas personas consideran inalcanzable, pero tenemos la suerte de ser un gran grupo, y nuestros esfuerzos comienzan a dar fruto.

\* \* \*

El contenido de este libro es ambicioso. Describe una teoría exhaustiva de cómo funciona el cerebro. Detalla qué es la inteligencia y cómo la crea tu cerebro. La teoría que presento no es completamente nueva. Muchas de las ideas particulares que está a punto de leer han existido antes de una forma u otra, pero no juntas de un modo coherente. Esto es algo que cabía esperar. Se dice que las "nuevas ideas" suelen ser viejas ideas reordenadas y reinterpretadas, lo que sin duda resulta aplicable a la teoría que se propone en este texto; pero el reordenamiento y la interpretación pueden suponer una diferencia abismal, la diferencia entre un cúmulo de detalles y una teoría satisfactoria. Espero que le cause la misma impresión que provoca a mucha gente. La reacción típica que escucho es: "Tiene sentido. No había pensado en la inteligencia de esa manera, pero ahora que me lo ha descrito puedo ver cómo encaja todo". Con este conocimiento la mayoría de las personas comienzan a contemplarse de una forma algo diferente. Empiezas a observar tu comportamiento diciendo: "Comprendo lo que acaba de suceder en mi cabeza". Espero que cuando haya terminado este libro tenga una nueva percepción de por qué piensa como piensa y se comporta como se comporta. También espero que algunos lectores se sientan inspirados para dirigir sus carreras a la construcción de máquinas inteligentes basadas en los principios esbozados en estas páginas.

Suelo referirme a esta teoría y mi planteamiento de estudio de la inteligencia como "inteligencia real" para distinguirlos de la "inteligencia artificial". Los científicos de la inteligencia artificial trataron de programar ordenadores para que actuaran como seres humanos sin resolver primero qué es la inteligencia y qué significa comprender. Dejaron fuera la parte más importante para construir máquinas inteligentes: la inteligencia. La "inteligencia real" establece la premisa de que antes de intentar construir máquinas inteligentes debemos comprender primero cómo piensa el cerebro, y no hay nada artificial en ello. Solo entonces podremos preguntarnos cómo construir máquinas inteligentes.

El libro se inicia con algunos antecedentes de por qué han fracasado los intentos previos de comprender la inteligencia y construir máquinas inteligentes. Luego presento y desarrollo la idea central de la teoría, lo que denomino el marco de la memoria-predicción. En el capítulo 6 detallo cómo el cerebro físico pone en práctica el modelo de memoria-predicción; en otras palabras, cómo funciona realmente el cerebro. A continuación explico las repercusiones sociales y demás de la teoría en el que para muchos lectores tal vez constituya el epígrafe que más reflexiones suscita. El libro concluye con un examen de las máquinas inteligentes: cómo podemos construirlas y cómo será el futuro. Espero que le resulte fascinante. Estas son algunas de las preguntas de las que nos ocuparemos a lo largo del texto:

### ***¿Pueden ser inteligentes los ordenadores?***

Durante décadas los científicos del campo de la inteligencia artificial han declarado que los ordenadores serán inteligentes cuando cuenten con la potencia necesaria. Yo no lo creo, y explicaré por qué. Los cerebros y los ordenadores hacen cosas radicalmente diferentes.

### ***¿No se suponía que las redes neuronales conducirían a las máquinas inteligentes?***

Sabemos que el cerebro está constituido por una red de neuronas, pero si primero no entendemos qué es lo que hace, unas simples redes neuronales no obtendrán mejores resultados para crear máquinas inteligentes que los programas informáticos.

### ***¿Por qué ha sido tan difícil entender cómo funciona el cerebro?***

La mayoría de los científicos afirman que debido a la complejidad del cerebro tardaremos mucho tiempo en comprenderlo. No estoy de acuerdo. La complejidad es un sinónimo de confusión, no una causa. Lo que yo sostengo es que unas cuantas asunciones intuitivas pero inexactas nos han inducido a error. El mayor es creer que la inteligencia se define por la conducta inteligente.

### ***¿Qué es la inteligencia si no se define por la conducta?***

El cerebro emplea enormes cantidades de memoria para crear un modelo del mundo. Todo lo que conoces y has aprendido se almacena en este modelo. El cerebro usa dicho modelo basado en la memoria para efectuar predicciones continuas sobre acontecimientos futuros. La capacidad para efectuar predicciones sobre el futuro constituye el quid de la inteligencia. Describiré la capacidad predictiva del cerebro a fondo, pues es la idea central de este libro.

### ***¿Cómo funciona el cerebro?***

La morada de la inteligencia es la corteza cerebral. Aunque posee un gran número de capacidades y una flexibilidad enorme, la corteza cerebral es sorprendentemente regular en sus detalles estructurales. Sus partes diferentes, sean responsables de la visión, el oído, el tacto o el lenguaje, funcionan según los mismos principios. La clave para entender la corteza cerebral es comprender estos principios comunes y, en particular, su estructura jerárquica. Examinaremos la corteza cerebral con el detalle suficiente para mostrar cómo capta su estructura la configuración del

mundo. Esta explicación será la parte más técnica del libro, pero los lectores interesados serán capaces de comprenderla aunque no sean científicos.

### ***¿Cuáles son las repercusiones de esta teoría?***

La teoría del cerebro puede ayudar a explicar cosas tales como nuestro modo de ser creativos, por qué tenemos conciencia de algo, por qué mostramos prejuicios, cómo aprendemos y por qué a los "perros viejos" les cuesta aprender "trucos nuevos". Analizaré varios de estos temas. En líneas generales, esta teoría nos proporciona una percepción de lo que somos y de por qué hacemos lo que hacemos.

### ***¿Podemos construir máquinas inteligentes? ¿Qué harán?***

Sí. Podemos y lo haremos. Considero que las capacidades de dichas máquinas evolucionarán con rapidez en unas cuantas décadas y en direcciones interesantes. Algunas personas temen que las máquinas inteligentes puedan resultar peligrosas para la humanidad, pero yo no comparto esa idea. Los robots no van a invadirnos. Por ejemplo, será mucho más fácil construir máquinas que aventajen nuestras facultades en pensamiento elevado de ámbitos como la física y la matemática que construir algo parecido a los robots parlantes que vemos en la ficción popular. Exploraré las increíbles direcciones en las que es probable que avance esta tecnología.

Mi meta es explicar esta nueva teoría de la inteligencia y del funcionamiento del cerebro de un modo que cualquiera sea capaz de entenderla. Una buena teoría debe resultar fácil de comprender, y no ocultarse en la jerga o en un argumento enrevesado. Comenzaré con un marco básico y luego iré añadiendo detalles a medida que avancemos. Algunos se razonarán atendiendo a la lógica; otros supondrán aspectos particulares del sistema de circuitos cerebral. Sin duda, algunos de los detalles de lo que propongo serán erróneos, lo que siempre sucede en cualquier campo de la ciencia. Tardará años en desarrollarse una teoría plenamente acabada, pero eso no disminuye la fuerza de la idea central.

\* \* \*

Cuando comencé a interesarme por los cerebros hace muchos años, fui a la biblioteca local para buscar un buen libro que explicara el funcionamiento cerebral. De adolescente me había acostumbrado a ser capaz de encontrar libros bien escritos que explicaban casi cualquier tema que me interesara. Los había sobre la teoría de la relatividad, los agujeros negros y la matemática, sobre cualquier cosa que me fascinara en ese momento. Sin embargo, mi búsqueda de un libro satisfactorio sobre el cerebro resultó vana. Acabé dándome cuenta de que nadie tenía idea de cómo funcionaba en realidad. Ni siquiera existían teorías malas o sin



verificar; simplemente no había ninguna, lo cual era algo poco habitual. Por ejemplo, por entonces nadie sabía cómo habían muerto los dinosaurios, pero había multitud de teorías, y de todas se podía aprender algo. Pero no era así en el caso de los cerebros. Al principio me costó creerlo. Me preocupaba que no supiéramos cómo funcionaba este órgano crucial. Mientras estudiaba lo que sí sabíamos sobre los cerebros, llegué a creer que debía haber una explicación sencilla. El cerebro no era mágico, y no me parecía que las respuestas fueran a ser tan complejas. El matemático Paul Erdos creía que las pruebas matemáticas más sencillas ya existían en un libro etéreo y que la labor del matemático era encontrarlas, "leer el libro". Del mismo modo, me parecía que la explicación de la inteligencia estaba "ahí fuera". Podía palparlo. Quería leer el libro.

Durante los últimos veinticinco años he tenido una visión de ese libro pequeño y claro sobre el cerebro. Era como una zanahoria que me ha mantenido motivado durante todo ese tiempo. Esta visión ha dado forma al libro que ahora tiene en sus manos. Nunca me ha gustado la complejidad, ni en la ciencia ni en la tecnología. Puede verlo reflejado en los productos que he diseñado, que se suelen señalar por su facilidad de uso. Las cosas más convincentes son sencillas. Así pues, este libro propone una teoría sencilla y directa sobre la inteligencia. Espero que le guste.

# 1

## Inteligencia Artificial

**Cuando** en junio de 1979 obtuve la licenciatura en ingeniería eléctrica en Cornell, no tenía ningún plan importante para mi vida. Empecé a trabajar como ingeniero en el nuevo campus de Intel en Portland (Oregón). La industria de microordenadores estaba comenzando, e Intel se hallaba en su centro. Mi labor consistía en analizar y resolver los problemas encontrados por los demás ingenieros que realizaban trabajo de campo con nuestro producto principal, los ordenadores de un solo sustrato de silicio. (Por aquel entonces, la reciente invención del microprocesador por parte de Intel había hecho posible que se pusiera un ordenador completo en una única placa de circuito integrado.) Publicaba un boletín de noticias, realizaba algunos viajes y tuve la oportunidad de conocer a los clientes. Era joven y me lo pasaba bien, aunque echaba de menos a mi novia de la universidad, que había aceptado un trabajo en Cincinnati.

Unos meses después me topé con algo que iba a cambiar la dirección de mi vida. Ese algo fue el recién publicado número de septiembre de *Scientific American*, que se dedicaba por completo al cerebro. Reavivó mi interés de la infancia por los cerebros y me resultó fascinante. Por él me enteré de la organización, desarrollo y química del cerebro, de los mecanismos neuronales de la visión, el movimiento y otras especializaciones, y de la base biológica de los trastornos mentales. Fue uno de los mejores números de *Scientific American* de todos los tiempos. Varios neurocientíficos con los que he hablado me han confirmado que desempeñó un papel significativo en la elección de su carrera, igual que me sucedió a mí.

El artículo final, "Thinking About the Brain", lo escribía Francis Crick, codescubridor de la estructura del ADN, que por entonces había dirigido su talento al estudio del cerebro. Crick sostenía que, a pesar de la constante acumulación de conocimiento detallado sobre el cerebro, seguía siendo un profundo misterio su funcionamiento. Los científicos no suelen escribir sobre lo que desconocen, pero a Crick no le importaba. Era como el niño señalando al emperador sin ropa. Según él, la neurociencia consistía en un montón de datos sin una teoría. Sus palabras exactas eran: "Lo que llama la atención es la falta de un marco amplio de ideas". A mi entender, era el educado modo británico de decir: "No tenemos ni una pista de cómo funciona esa cosa". Era cierto entonces y sigue siéndolo en la actualidad.

Las palabras de Crick fueron como una llamada para mí. El deseo de toda mi vida de comprender los cerebros y construir máquinas inteligentes cobró vida. Aunque apenas acababa de abandonar la universidad, decidí cambiar de carrera. Iba a estudiar los cerebros, no solo a entender cómo funcionan, sino a emplear ese conocimiento como base para nuevas tecnologías, para construir máquinas inteligentes. Tardaría algún tiempo en poner este plan en práctica.

En la primavera de 1980 me trasladé a la sede de Intel en Boston para reunirme con mi futura esposa, que iba a empezar un curso de posgrado. Ocupé un puesto para enseñar a los clientes y empleados a diseñar sistemas basados en los microprocesadores. Pero tenía la vista puesta en una meta diferente: trataba de resolver cómo trabajar en la teoría del cerebro. El ingeniero que había en mí se daba cuenta de que una vez que comprendiéramos cómo funcionan los cerebros podríamos construirlos, y el modo natural de construir cerebros artificiales era en silicio. Trabajaba para la compañía que inventó el chip de memoria y el microprocesador de silicio, así que tal vez fuera posible interesar a Intel para que me permitiera dedicar parte de mi tiempo a pensar sobre la inteligencia y cómo diseñar chips de memoria semejantes al cerebro. Escribí una carta al director de Intel, Gordon Moore, que puede resumirse en la siguiente:

Estimado doctor Moore:

Propongo que iniciemos un grupo de investigación que se dedique a comprender cómo funciona el cerebro. Puede empezar con una persona, yo, y avanzar desde ahí. Tengo plena confianza en que somos capaces de entenderlo. Será una gran empresa algún día.

—Jeff Hawkins

Moore me puso en contacto con el principal científico de Intel, Ted Hoff. Tomé un avión a California para reunirme con él y exponer mi propuesta para estudiar el cerebro. Hoff era famoso por dos cosas. La primera —de la que yo estaba bien enterado—, su labor en el diseño del primer microprocesador. La segunda —que yo no conocía por aquel entonces— su trabajo en la incipiente teoría de redes neuronales. Tenía experiencia con neuronas artificiales y algunas de las cosas que se podían hacer con ellas. Yo no estaba preparado para esto. Tras escuchar mi propuesta, dijo que no le parecía posible resolver cómo funciona el cerebro en el futuro previsible y, por lo tanto, no tenía sentido que Intel me apoyara. Hoff estaba en lo cierto, pues ahora, veinticinco años después, estamos empezando a efectuar un avance significativo en la comprensión de los cerebros. La oportunidad lo es todo en los negocios. No obstante, en ese momento sentí una gran desilusión.

Tiendo a buscar el camino con menores obstáculos para alcanzar mis metas. Trabajar en los cerebros en Intel habría sido la transición más sencilla. Una vez eliminada esa opción, indagué la alternativa siguiente. Decidí presentarme para un curso de posgrado en el Massachusetts Institute of Technology, que era famoso por su investigación sobre la inteligencia artificial y estaba bien comunicado. Parecía encajar bien. Yo poseía una amplia formación en ciencia informática —“bien”—. Deseaba construir máquinas inteligentes —“bien”—. Quería estudiar primero los cerebros para ver cómo funcionan —“vaya, eso es un problema”—. Esta última meta, querer comprender cómo funcionaban los cerebros, constituía un

problema a los ojos de los científicos del laboratorio de inteligencia artificial del MIT.

Fue como toparme con un muro de ladrillos. El MIT era el buque insignia de la inteligencia artificial. En la época en que solicité el ingreso en el MIT, albergaba a docenas de personas brillantes cautivadas con la idea de programar ordenadores para producir una conducta inteligente. Para estos científicos, la visión, el lenguaje, la robótica y la matemática no eran más que problemas de programación. Los ordenadores podían hacer cualquier cosa que hiciera un cerebro y más, así que ¿por qué constreñir el pensamiento con el desorden biológico del ordenador de la Naturaleza? Estudiar los cerebros limitaría la reflexión. Creían que era mejor estudiar los límites últimos de la informática, cuya máxima expresión eran los ordenadores digitales. Su santo grial era crear programas informáticos que igualaran, primero, y sobrepasaran, después, las facultades humanas. Adoptaban el planteamiento de que el fin justifica los medios; no les interesaba cómo funcionaban los cerebros reales. Algunos se enorgullecían de ignorar la neurobiología.

Me pareció una forma equivocada de abordar el problema. Intuía que el planteamiento de la inteligencia artificial no solo fracasaría en crear programas que hicieran lo que saben hacer los humanos, sino que no nos enseñaría qué era la inteligencia. Los ordenadores y los cerebros están contruidos según principios completamente diferentes. Uno está programado; el otro aprende por sí mismo. Uno tiene que trabajar a la perfección; el otro es flexible por naturaleza y tolera los fallos. Uno cuenta con un procesador central; el otro carece de control centralizado. Y la lista de diferencias prosigue y prosigue. La principal razón por la que pensaba que los ordenadores no serían inteligentes era que comprendía cómo funcionan hasta el nivel de la física de los transistores, y este conocimiento me aportaba la sensación intuitiva de que los cerebros y los ordenadores eran esencialmente distintos. No podía probarlo, pero lo sabía en la medida en que se puede saber algo por intuición. Tal vez a la larga, razonaba, la inteligencia artificial lleve a la invención de productos útiles, pero no iba a construir máquinas verdaderamente inteligentes.

En cambio, yo quería comprender la inteligencia y percepción reales, estudiar la psicología y la anatomía del cerebro, enfrentarme al reto de Francis Crick y presentar un amplio marco del funcionamiento cerebral. En particular, tenía puesta la vista en la corteza cerebral, la parte del cerebro de los mamíferos cuyo desarrollo era más reciente y donde moraba la inteligencia. Una vez que comprendiéramos cómo funcionaba la corteza cerebral, podríamos acometer la construcción de máquinas inteligentes, pero no antes.

Por desgracia, los profesores y alumnos que conocí en el MIT no compartían mis intereses. No creían que se necesitara estudiar cerebros reales para comprender la

inteligencia y construir máquinas inteligentes. Así me lo dijeron. En 1981 la universidad rechazó mi solicitud de ingreso.

\* \* \*

**M**ucha gente cree en la actualidad que la inteligencia artificial está sana y salva, y que solo espera que los ordenadores tengan la potencia suficiente para hacer realidad sus muchas promesas. Cuando los ordenadores tengan suficiente memoria y potencia de procesamiento —prosigue esta creencia—, los programadores de la inteligencia artificial serán capaces de crear máquinas inteligentes. No estoy de acuerdo. La inteligencia artificial presenta un defecto fundamental: no aborda de forma adecuada qué es la inteligencia o qué significa entender algo. Una breve ojeada a su historia y los principios sobre los que se edificó ayudará a explicar cómo la disciplina se ha desviado de su rumbo.

La propuesta de la inteligencia artificial nació con el ordenador digital. Una figura clave al inicio de este movimiento fue el matemático inglés Alan Turing, uno de los inventores de la idea del ordenador de uso general. Su golpe maestro fue demostrar formalmente el concepto de cómputo universal; esto es, que todos los ordenadores son equivalentes en lo básico, prescindiendo de los detalles de su construcción. Como parte de su demostración, concibió una máquina imaginaria con tres partes esenciales: una caja de procesamiento, una cinta de papel y un aparato que lee y escribe marcas en la cinta mientras se mueve de un lado a otro. La cinta servía para guardar información, como el famoso código informático de unos y ceros (fue antes de la invención de los chips de memoria o la unidad de disco, así que Turing ideó la cinta de papel para el almacenamiento). La caja, que en la actualidad denominamos unidad de procesamiento central (CPU), sigue un conjunto de reglas fijas para leer y editar la información de la cinta. Turing demostró matemáticamente que si se elige el conjunto de reglas preciso para la CPU y se le proporciona una cinta indefinidamente larga con la que trabajar puede realizar cualquier conjunto de operaciones del Universo. Sería una de las muchas máquinas equivalentes que ahora denominamos Máquinas de Turing Universales. Ya se trate de calcular raíces cuadradas o trayectorias de balas, jugar a un juego, editar fotos o cuadrar transacciones bancadas, por debajo no hay más que unos y ceros, y cualquier Máquina de Turing podría programarse para efectuar la tarea. El procesamiento de la información es procesamiento de la información... Todos los ordenadores digitales son equivalentes en su lógica.

La conclusión de Turing era indiscutiblemente cierta y muy provechosa. La revolución informática y todos sus productos se basan en ella. Después, Turing pasó a la cuestión de cómo construir una máquina inteligente. Le parecía que los ordenadores podían ser inteligentes, pero no quería entrar en discusiones sobre si era posible o no. Tampoco se consideraba capaz de definir la *inteligencia* formalmente, así que no lo intentó. En su lugar, propuso una prueba de existencia

para la inteligencia, el famoso Test de Turing: si un ordenador puede engañar a un interrogador humano y lograr que piense que él también es una persona, el ordenador tiene que ser inteligente por definición. De este modo, con su test como vara de medir y su máquina como medio, Turing ayudó a lanzar la disciplina de la inteligencia artificial. Su dogma central: el cerebro no es más que otro tipo de ordenador. Carece de importancia cómo se diseñe un sistema artificialmente inteligente; solo tiene que producir conducta semejante a la humana.

Los defensores de la inteligencia artificial veían paralelos entre los ordenadores y el pensamiento. Decían: "Veamos, las proezas más impresionantes de la inteligencia humana suponen, sin duda, la manipulación de símbolos abstractos, y eso es lo que hacen también los ordenadores. ¿Qué hacemos cuando hablamos o escuchamos? Manipulamos símbolos mentales denominados palabras, utilizando reglas gramaticales bien definidas. ¿Qué hacemos cuando jugamos al ajedrez? Empleamos símbolos mentales que representan las propiedades y situación de las diversas piezas. ¿Qué hacemos cuando vemos? Usamos símbolos mentales para representar objetos, sus posiciones, sus nombres y otras propiedades. Sin duda, la gente hace todo esto con los cerebros y no con la clase de ordenadores que construimos, pero Turing ha demostrado que no importa cómo se pongan en ejecución o se manipulen los símbolos. Se puede hacer con una reunión de dientes y engranajes, con un sistema de interruptores eléctricos o con la red de neuronas del cerebro, siempre que el medio empleado pueda percibir la equivalencia funcional de una Máquina de Turing Universal".

Esta asunción se vio reforzada por un influyente artículo científico publicado en 1943 por el neuropsicólogo Warren McCulloch y el matemático Walter Pitts. Describían cómo las neuronas podían realizar funciones digitales; es decir, cómo cabía la posibilidad de que las células nerviosas pudieran reproducir la lógica formal que constituye el núcleo de los ordenadores. La idea era que las neuronas podían actuar como las que los ingenieros denominan puertas lógicas. Estas ejecutan operaciones lógicas sencillas como Y, NO y O. Los chips de los ordenadores están compuestos por millones de puertas lógicas, conectadas todas en circuitos precisos y complejos. Una CPU no es más que un grupo de puertas lógicas.

McCulloch y Pitts señalaron que las neuronas también podían conectarse de formas precisas para realizar funciones lógicas. Como las neuronas reúnen aportaciones unas de otras y las procesan para decidir si lanzan un resultado, era concebible que pudieran ser puertas lógicas vivas. De este modo, deducían, cabía la posibilidad de que el cerebro estuviera construido con puertas Y, puertas O y otros elementos lógicos creados con neuronas, en analogía directa con el cableado de los circuitos electrónicos digitales. No está claro si McCulloch y Pitts creían de verdad que el cerebro funcionaba de este modo; solo decían que era posible. Y desde la lógica esta visión de las neuronas es posible. En teoría, las neuronas

pueden ejecutar funciones digitales. Sin embargo, nadie se molestó en preguntarse si esta era la forma real en la que se conectaban las neuronas en el cerebro. Lo tomaron como demostración, sin tener en cuenta la falta de pruebas biológicas de que los cerebros no eran más que otra clase de ordenador.

También conviene señalar que la filosofía de la inteligencia artificial se vio respaldada por la tendencia que dominó la psicología durante la primera mitad del siglo XX, llamada conductismo. Los conductistas creían que no era posible saber qué sucede dentro del cerebro, al cual denominaban la caja negra impenetrable. Pero se podía observar y calibrar el entorno de un animal y sus conductas, lo que siente y lo que hace. Reconocían que el cerebro contenía mecanismos reflejos que podían usarse para condicionar a un animal a fin de que adoptara nuevas conductas mediante la recompensa y los castigos. Pero, aparte de esto, sostenían que para entender algo no se necesitaba estudiar el cerebro, ni sentimientos tan complicados y subjetivos como el hambre, el miedo, o lo que significan. Huelga decir que esta filosofía de la investigación acabó marchitándose durante la segunda mitad del siglo XX, pero la inteligencia artificial seguiría vigente mucho más tiempo.

Cuando terminó la Segunda Guerra Mundial y se dispuso de ordenadores digitales para aplicaciones más amplias, los pioneros de la inteligencia artificial se remangaron y comenzaron a programar. ¿Traducción de lenguas? ¡Fácil! No hay más que descifrar un código. Solo necesitamos colocar cada símbolo del Sistema A en su semejante del Sistema B. ¿Visión? Eso también parece fácil. Ya conocemos los teoremas geométricos que tratan de la rotación, la escala y el desplazamiento, y será sencillo codificarlos como algoritmos informáticos, así que en esto tenemos ya medio camino adelantado. Los expertos en inteligencia artificial realizaron grandes declaraciones sobre la rapidez con la que la inteligencia de los ordenadores igualaría, primero, y sobrepasaría, después, la humana.

Resulta irónico que el programa informático que estuvo más cerca de superar el test de Turing, un programa llamado Eliza, imitara a un psicoanalista, reformulando las preguntas que se le hacían. Por ejemplo, si una persona tecleaba: "Mi novio y yo ya no nos hablamos", Eliza diría: "Cuéntame más de tu novio" o "¿Por qué piensas que tu novio y tú ya no os habláis?". Diseñado como una broma, el programa llegó a engañar a algunos, aunque era tonto y superficial. Entre los esfuerzos más serios se incluyeron programas como Blocks World, una habitación simulada que contenía bloques de diferentes colores y formas. Se podían plantear preguntas a Blocks World del tipo: "¿Hay una pirámide verde sobre el cubo rojo grande?", o solicitar: "Pon el cubo azul sobre el cubo rojo pequeño". El programa respondía la pregunta o trataba de hacer lo que se le pedía. Todo era simulado, y funcionaba. Pero se limitaba a su pequeño mundo artificial de bloques. Los programadores no fueron capaces de generalizarlo para que hiciera algo útil.

Mientras tanto, el público vivía impresionado por la sucesión continua de éxitos aparentes y nuevas historias sobre la tecnología de la inteligencia artificial. Un programa que generó entusiasmo inicial era capaz de resolver teoremas matemáticos. Desde Platón, la inferencia deductiva de múltiples pasos se venía considerando la culminación de la inteligencia humana, así que al principio pareció que la inteligencia artificial había conseguido el premio gordo. Pero, como en el caso del Blocks World, resultó que el programa solo podía hallar teoremas muy sencillos que ya se conocían. Luego se suscitó una gran agitación con los "sistemas expertos", bases de datos que podían responder preguntas planteadas por los usuarios humanos. Por ejemplo, un sistema experto médico sería capaz de diagnosticar la enfermedad de un paciente si se le aportaba una lista de síntomas. Pero de nuevo resultaron ser de uso limitado y no mostraron nada que se aproximara a una inteligencia generalizada. Los ordenadores podían jugar al ajedrez con pericia de expertos, y el *Deep Blue* de IBM se hizo famoso por acabar derrotando a Gary Kasparov, el campeón mundial, en su propio juego. Pero estos éxitos fueron vanos. *Deep Blue* no ganó por ser más inteligente que un ser humano, sino porque era millones de veces más rápido. *Deep Blue* no tenía intuición. Un jugador experto humano mira una posición del tablero y de inmediato ve qué zonas de juego tienen mayores posibilidades de ser provechosas o peligrosas, mientras que un ordenador no posee un sentido innato de lo que es importante y debe explorar muchas más opciones. *Deep Blue* tampoco tenía sentido de la historia del juego y no sabía nada de su rival. Jugaba al ajedrez, mas no lo entendía, del mismo modo que una calculadora realiza operaciones aritméticas, pero no entiende la matemática.

En todos los casos, los programas de inteligencia artificial que lograban éxito solo servían para efectuar la tarea particular especificada en su diseño. No generalizaban ni mostraban flexibilidad, e incluso sus creadores admitían que no pensaban como seres humanos. Algunos de los problemas de la inteligencia artificial que al principio parecieron fáciles no consiguieron ser resueltos. En la actualidad, ningún ordenador puede comprender el lenguaje tan bien como un niño de tres años ni ver tan bien como un ratón.

Tras muchos años de esfuerzo, promesas incumplidas y ningún éxito rotundo, la inteligencia artificial comenzó a perder su lustre. Los científicos de la disciplina pasaron a otros campos de investigación. Las empresas dedicadas a ella fracasaron y la financiación se hizo más escasa. Programar ordenadores para que realizaran las tareas más básicas de percepción, lenguaje y conducta comenzó a parecer imposible. Hoy la situación no ha cambiado mucho. Como ya he afirmado, todavía hay personas que creen que los problemas de la inteligencia artificial pueden resolverse con ordenadores más rápidos, pero la mayoría de los científicos piensan que la empresa en sí era defectuosa.



No debemos culpar a los pioneros de la inteligencia artificial por sus fracasos. Alan Turing fue brillante. Todos podían decir que la Máquina de Turing cambiaría el mundo; y lo hizo, pero no mediante la inteligencia artificial.

\* \* \*

**Mi** escepticismo hacia las afirmaciones sobre la inteligencia artificial se agrandó por la misma época en que solicité mi ingreso en el MIT. John Searle, influyente profesor de filosofía de la Universidad de California, en Berkeley, se dedicaba por aquel entonces a sostener que los ordenadores no eran y no podían ser inteligentes. Para demostrarlo, en 1980 presentó un experimento mental denominado la Habitación China. Consiste en lo siguiente:

Supongamos que hay una habitación con una ranura en una pared, y dentro se encuentra una persona de lengua inglesa sentada ante un escritorio. Tiene un gran libro de instrucciones y todos los lápices y papel de borrador que pueda necesitar. Hojeando el libro, ve que las instrucciones, escritas en inglés, dictan modos de manipular, clasificar y comparar los caracteres chinos. Pero, ojo, las instrucciones no dicen nada sobre el significado de los caracteres chinos; solo tratan de cómo deben copiarse, borrarse, reordenarse, transcribirse y demás.

Alguien desde fuera de la habitación desliza un trozo de papel por la ranura. En él está escrito un relato y preguntas sobre dicho relato, todo en chino. El hombre de dentro no habla ni lee una palabra de chino, pero recoge el papel y prosigue trabajando con el libro de instrucciones. A veces estas le indican que escriba caracteres en el papel de borrador y otras que cambie y borre caracteres. Aplicando una regla tras otra, escribiendo y borrando caracteres, el hombre se afana hasta que el libro le anuncia que ha terminado. Cuando por fin ha acabado, ha escrito una nueva página de caracteres que sin él saberlo son las respuestas a las preguntas. El libro le indica que pase su papel por la ranura. Lo hace y se pregunta de qué iba todo este tedioso ejercicio.

Fuera, una persona que habla chino lee la página. Todas las respuestas son correctas, señala; incluso perspicaces. Si se le preguntara si esas respuestas provenían de una mente inteligente que había comprendido el relato, diría sin duda que sí. ¿Pero está en lo cierto? ¿Quién entendió el relato? No fue la persona de dentro, por supuesto. No fue el libro, que después de todo no es más que un volumen colocado inerte sobre el escritorio entre montones de papel. Así pues, ¿cuándo se efectuó la comprensión? La respuesta de Searle es que no hubo comprensión; solo se hojearon muchas páginas sin sentido y se copió a lápiz y borró sin saber lo que se hacía. Y ahora viene lo bueno: la habitación china es análoga a un ordenador digital. La persona es la CPU, que ejecuta instrucciones sin pensar, y el papel de borrador es la memoria. De este modo, por mucho ingenio que se ponga en el diseño de un ordenador para simular inteligencia y producir la misma conducta de un ser humano, dicho ordenador no posee entendimiento y no

es inteligente. (Searle dejó claro que no sabía qué es la inteligencia; se limitaba a afirmar que, fuera lo que fuese, los ordenadores no la tenían.)

Este razonamiento creó una enorme disputa entre filósofos y expertos en inteligencia artificial. Generó cientos de artículos, además de algo de sarcasmo y mala sangre. Los defensores de la inteligencia artificial contraatacaron a Searle con docenas de argumentos, declarando, entre otras cosas, que si bien ninguna de las partes componentes de la habitación entendía chino, la habitación como un todo sí lo entendía, o que la persona de la habitación sí lo entendía, aunque no lo sabía. Yo pienso que Searle estaba en lo cierto. Cuando reflexioné sobre el razonamiento de la habitación china y sobre el funcionamiento de los ordenadores, no vi que hubiera entendimiento en ningún lugar. Llegué a la convicción de que necesitábamos comprender qué es el "entendimiento", un modo de definirlo que pusiera de manifiesto cuándo un sistema era inteligente y cuándo no, cuándo entiende chino y cuándo no. Su conducta no nos lo dice.

Un ser humano no necesita "hacer" nada para entender un relato. Puedo leer un relato en silencio, y aunque no manifieste una conducta evidente, mi entendimiento y comprensión son claros, al menos para mí. Por otro lado, usted no es capaz de afirmar por mi conducta callada si he entendido o no el relato, ni siquiera si conozco la lengua en la que está escrito. Podría plantearme después preguntas para comprobarlo, pero mi entendimiento ocurrió cuando leí el relato, no cuando respondo a sus preguntas. Este libro sostiene la tesis de que el entendimiento no puede medirse por la conducta externa; como veremos en los capítulos siguientes; es más bien una medición interna de cómo recuerda las cosas el cerebro y emplea su memoria para hacer predicciones. La habitación china, *Deep Blue* y la mayoría de los programas informáticos no poseen nada semejante. No comprenden lo que están haciendo. El único modo de juzgar si un ordenador es inteligente es valorar su información de salida, o conducta.

El argumento defensivo supremo de la inteligencia artificial es que los ordenadores, en teoría, podrían simular el cerebro completo. Un ordenador podría copiar el modelo de todas las neuronas y sus conexiones, en cuyo caso no habría nada que distinguiera la "inteligencia" del cerebro de la "inteligencia" de la simulación por ordenador. Aunque tal vez resulte imposible de llevar a cabo en la práctica, estoy de acuerdo. Pero los investigadores de la inteligencia artificial no simulan cerebros, y sus programas no son inteligentes. No se puede simular un cerebro si primero no se comprende qué hace.

\* \* \*

**D**espués de que me rechazaran tanto Intel como el MIT, no sabía qué hacer. Y cuando no sabes cómo actuar, la mejor estrategia suele ser no efectuar cambios hasta que se aclaren las opciones. Así que continué trabajando en el campo de la informática. Me encontraba a gusto en Boston, pero en 1982 mi esposa quiso que

nos trasladáramos a California, y así lo hicimos (era, una vez más, el camino con menos obstáculos). Encontré trabajo en Silicon Valley, en una empresa informática llamada Grid Systems. Dicha empresa inventó el ordenador portátil, una bonita máquina que se convirtió en el primer ordenador de la colección del Museo de Arte Moderno de Nueva York. Trabajé primero en mercadotecnia y luego en ingeniería, y acabé creando un lenguaje de programación de alto nivel llamado GridTask. Él y yo nos fuimos haciendo cada vez más importantes para el éxito de Grid; mi carrera marchaba bien.

No obstante, no podía sacarme de la cabeza la curiosidad por el cerebro y las máquinas inteligentes. Me consumía el deseo de estudiar los cerebros. Así que hice un curso por correspondencia de psicología humana y estudié por mi cuenta (nadie había sido rechazado jamás por una escuela de correspondencia). Tras aprender una buena cantidad de biología, decidí solicitar el ingreso en un curso de posgrado de un programa de biología y estudiar la inteligencia desde dentro de las ciencias biológicas. Si la ciencia informática no quería un teórico del cerebro, puede que el mundo de la biología aceptara a un científico de los ordenadores. Por aquel entonces no existía biología teórica ni nada semejante, y mucho menos neurociencia teórica, por lo cual la biofísica parecía el mejor campo para mis intereses. Estudié mucho, pasé los exámenes de ingreso requeridos, preparé un currículo, pedí cartas de recomendación y conseguí ser aceptado como alumno de posgrado a tiempo completo en el programa de biofísica de la Universidad de California, en Berkeley.

Estaba emocionado. Por fin podía comenzar en serio con la teoría sobre el cerebro, o eso pensaba. Dejé mi puesto en Grid sin intención de volver a trabajar en la industria informática. Por supuesto, ello significaba renunciar indefinidamente a mi sueldo. Mi esposa creía que había llegado el momento de comprar una casa y formar una familia, y yo disponía alegremente de dejar de aportar ingresos. Este no era en modo alguno el camino con menores obstáculos, pero era la mejor opción que tenía, y mi esposa apoyó mi decisión.

John Ellenby, fundador de Grid, me metió en su despacho justo antes de que me marchara y me dijo: "Sé que esperas no volver jamás a Grid ni a la industria informática, pero nunca se sabe lo que va a pasar. En lugar de dejar el puesto por completo, ¿por qué no pides una excedencia? De este modo, si regresas en un año o dos, puedes retomar tu sueldo, puesto y opción sobre acciones donde los dejaste". Fue un bonito gesto, y lo acepté; pero sentía que estaba abandonando el mundo de la informática para siempre.

## 2

### Redes Neuronales

**Cuando** ingresé en la Universidad de Berkeley, en enero de 1986, lo primero que hice fue compilar una historia de las teorías sobre la inteligencia y el funcionamiento cerebral. Leí cientos de artículos de anatomistas, fisiólogos, filósofos, lingüistas, científicos informáticos y psicólogos. Numerosas personas desde muchos campos habían escrito extensamente sobre el pensamiento y la inteligencia. Cada campo contaba con su conjunto de publicaciones y cada uno empleaba su terminología propia. Sus descripciones me resultaron contradictorias e incompletas. Los lingüistas hablaban de la inteligencia con términos como "sintaxis" y "semántica". Para ellos el cerebro y la inteligencia solo trataban del lenguaje. Los científicos de la visión hacían referencia a esbozos en 2D, 2 1/2 D y 3D. Para ellos el cerebro y la inteligencia consistían en el reconocimiento de patrones visuales. Los científicos informáticos hablaban de esquemas y marcos, nuevos términos que habían inventado para representar el conocimiento. Ninguna de estas personas se ocupaba de la estructura del cerebro y de cómo llevaría a la práctica cualquiera de sus teorías. Por su parte, los anatomistas y neurofisiólogos escribían en abundancia sobre la estructura del cerebro y el comportamiento de las neuronas, pero en su mayoría evitaban cualquier intento de formular una teoría a gran escala. Era difícil y frustrante tratar de encontrar sentido a estos planteamientos diversos y la montaña de datos experimentales que los acompañaban.

Por esa época hizo su aparición en escena un planteamiento sobre las máquinas inteligentes novedoso y prometedor. Las redes neuronales estaban presentes de una forma u otra desde finales de la década de 1960, pero rivalizaban con el movimiento de la inteligencia artificial por el dinero y la atención de los organismos que financiaban la investigación. Durante varios años, los investigadores de las redes neuronales estuvieron en una lista negra y no conseguían fondos. Sin embargo, unas cuantas personas continuaron dedicándose a ellas, y a mediados de la década de 1980 llegó por fin su gran momento. Es difícil saber con exactitud por qué hubo un interés repentino por las redes neuronales, pero sin duda un factor decisivo fue el fracaso continuado de la inteligencia artificial. La gente buscaba alternativas y encontró una en ellas.

Las redes neuronales suponían una verdadera mejora sobre el planteamiento de la inteligencia artificial porque su arquitectura se basa, si bien en líneas muy generales, en sistemas nerviosos reales. En lugar de programar ordenadores, a los investigadores de las redes neuronales, también conocidos como *conectistas*, les interesaba aprender qué clases de conductas podían mostrarse enganchando juntas un montón de neuronas. Los cerebros están compuestos por neuronas; por lo tanto, el cerebro es una red neuronal. Eso es un hecho. Los conectistas tenían la esperanza de que las propiedades esquivas de la inteligencia se aclararían

estudiando cómo interactúan las neuronas, y de que algunos de los problemas insolubles de la inteligencia artificial pudieran resolverse reproduciendo las conexiones precisas entre poblaciones de neuronas. Una red neuronal se diferencia de un ordenador porque no tiene CPU y no guarda información en una memoria centralizada. El conocimiento y las memorias de la red se distribuyen por toda su conectividad, igual que en el caso de los cerebros reales.

A primera vista, las redes neuronales parecían encajar bien con mis intereses, pero ese campo acabó desilusionándome pronto. Para entonces ya me había formado la opinión de que había tres cosas que resultaban esenciales para comprender el cerebro. Mi primer criterio era la inclusión del tiempo en la función cerebral. Los cerebros reales procesan con rapidez cambiando flujos de información. No hay nada estático en el flujo de información que entra y sale del cerebro.

El segundo criterio era la importancia de la realimentación. Los neuroanatomistas sabían desde hacía mucho tiempo que el cerebro está saturado de conexiones de realimentación. Por ejemplo, en el circuito entre la corteza cerebral y una estructura inferior llamada tálamo, las conexiones hacia atrás (hacia la entrada) sobrepasan a las que van hacia delante casi diez veces, lo que quiere decir que por cada fibra que nutre de información hacia delante en la corteza cerebral hay diez fibras que nutren de información hacia atrás a los sentidos. La realimentación domina también la mayoría de las conexiones por toda la corteza cerebral. Nadie comprende el papel preciso de esta realimentación, pero por la investigación publicada resultaba evidente que existía en todas partes. Yo me imaginaba que debía de ser importante.

El tercer criterio era que toda teoría o modelo del cerebro debía explicar su arquitectura física. La corteza cerebral no es una estructura sencilla. Como veremos más adelante, está organizada en una jerarquía que se repite. Sin duda, cualquier red neuronal que no reconociera dicha estructura no iba a actuar como un cerebro.

Pero cuando el fenómeno de las redes neuronales ocupó la escena, se basó en su mayoría en una clase de modelos ultrasencillos que no cumplían ninguno de estos criterios. La mayor parte de las redes neuronales constaban de un pequeño número de neuronas conectadas entre sí en tres filas. Se presenta un modelo (la entrada) a la primera fila. Estas neuronas de entrada están conectadas con la fila siguiente, las denominadas unidades ocultas. Luego estas conectan con la fila final de neuronas, las unidades de salida. Las conexiones entre neuronas presentan fuerzas variables, lo que significa que la actividad en una neurona podría aumentar la actividad en otra y disminuirla en una tercera, dependiendo de las fuerzas de conexión. Cambiando dichas fuerzas, la red aprende a representar modelos de entrada en modelos de salida.

Estas redes neuronales sencillas solo procesaban modelos estáticos, no usaban realimentación y no se parecían en absoluto a cerebros. El tipo más común, denominado red de "propagación recurrente", aprendía transmitiendo un error desde las unidades de salida de nuevo hacia las unidades de entrada. Tal vez piensen que esta es una forma de realimentación, pero no es así. La propagación de errores hacia atrás solo ocurría durante la fase de aprendizaje. Cuando la red neuronal funcionaba con normalidad una vez que se había entrenado; la información solo fluía en un sentido. No había realimentación de las salidas a las entradas. Y los modelos no tenían sentido del tiempo. Un patrón de entrada estático se transformaba en un patrón de salida estático. Luego se presentaba otro patrón de entrada. No había historia o registro en la red de lo que había sucedido ni siquiera un poco antes. Y, por último, la arquitectura de estas redes neuronales era trivial comparada con la estructura compleja y jerárquica del cerebro.

Pensaba que la disciplina pasaría enseguida a ocuparse de redes más realistas, pero no lo hizo. Puesto que estas redes neuronales sencillas eran capaces de ejecutar cosas interesantes, la investigación pareció detenerse ahí durante años. Habían encontrado una herramienta nueva y atractiva, y de la noche a la mañana miles de científicos, ingenieros y estudiantes obtenían becas, se doctoraban y escribían libros sobre las redes neuronales. Se formaron empresas cuyo fin era emplear dichas redes para pronosticar los movimientos de la Bolsa, procesar las solicitudes de créditos, verificar firmas y realizar cientos de diversas aplicaciones de clasificación de patrones. Aunque la intención de los fundadores del campo tal vez fuera más general, acabó dominado por personas a las que no interesaba la comprensión del funcionamiento cerebral o en qué consiste la inteligencia.

La prensa popular no entendía bien dicha distinción. Los periódicos, revistas y programas de ciencia televisivos presentaban las redes neuronales como si fueran "semejantes al cerebro" o funcionaran "según los mismos principios que el cerebro". A diferencia de la inteligencia artificial, donde todo tenía que ser programado, las redes neuronales aprendían por el ejemplo, lo que resultaba de algún modo más inteligente. Una muestra destacada era NetTalk. Esta red neuronal aprendió a representar sucesiones de letras sobre sonidos hablados. Cuando se ejercitó a la red en texto impreso, comenzó a sonar como la voz de un ordenador leyendo las palabras. Era fácil imaginar que con algo más de tiempo las redes neuronales conversarían con los seres humanos. NetTalk fue anunciada erróneamente en las noticias nacionales como una máquina que aprendía a leer. Era una gran exhibición, pero lo que hacía en realidad rayaba en lo trivial. No leía, no entendía y no tenía ningún valor práctico. Se limitaba a casar combinaciones de letras con patrones de sonido predefinidos.

Permítanme citar una analogía para mostrar lo lejos que se hallaban las redes neuronales de los cerebros reales. Imaginemos que en lugar de intentar discurrir cómo lo hace un cerebro estuviéramos tratando de deducir cómo funciona un

ordenador digital. Tras años de estudio, descubrimos que todo en el ordenador está compuesto de transistores. Hay cientos de millones de transistores en un ordenador y están conectados entre sí de formas precisas y complejas. Pero no comprendemos cómo funciona dicho ordenador o por qué los transistores están conectados de ese modo. Por lo tanto, un día decidimos conectar solo unos cuantos para ver qué sucede. Y, quién lo iba a decir, descubrimos que cuando solo se conectan juntos tres transistores de una forma determinada se convierten en un amplificador. Una pequeña señal colocada en un extremo se amplifica en el otro. (Los amplificadores de las radios y televisiones se fabrican realmente empleando transistores de este modo.) Se trata de un descubrimiento importante, y de la noche a la mañana surge una industria que se dedica a fabricar radios, televisiones y otros aparatos de transistores utilizando dichos amplificadores. Todo ello está muy bien, pero no nos dice nada de cómo funciona el ordenador. Aunque el amplificador y el ordenador están hechos de transistores, no tienen casi nada más en común. Del mismo modo, un cerebro real y la red neuronal de tres filas se construyen con neuronas, pero no poseen casi nada más en común.

Durante el verano de 1987 tuve una experiencia que arrojó más agua fría sobre mi ya escaso entusiasmo por las redes neuronales. Asistí a una conferencia sobre el tema en la que vi una presentación de una empresa llamada Nestor. Dicha empresa trataba de vender la aplicación de una red neuronal para reconocer escritura manual sobre una tablilla. Ofrecía la licencia del programa por un millón de dólares, lo cual me llamó la atención. Aunque Nestor promocionaba la sofisticación de su algoritmo de red neuronal y lo vendía como otro importante avance, me pareció que el problema del reconocimiento de la letra podía resolverse de una forma más sencilla y tradicional. Esa noche volví a casa pensando en la cuestión, y en dos días había diseñado un reconocedor de letra más rápido, pequeño y flexible. No empleaba una red neuronal y no funcionaba en absoluto como un cerebro, si bien el algoritmo se inspiraba en una matemática que yo estaba estudiando en relación con los cerebros. Esa conferencia despertó mi interés por diseñar ordenadores con una interfaz de aguja (que acabó conduciendo a la PalmPilot diez años después). El reconocedor de letra que había creado se convirtió en la base del sistema de entrada de texto llamado Graffiti, empleado en la primera serie de productos de Palm. Creo que Nestor cerró.

Era demasiado para redes neuronales tan sencillas. La mayor parte de sus capacidades las proporcionaban fácilmente otros métodos, y el bombo y platillo con los que las habían recibido los medios de comunicación acabó sosegándose. Por lo menos, los investigadores de las redes neuronales no declaraban que sus modelos eran inteligentes. Después de todo, se trataba de redes extremadamente simples y hacían menos que los programas de inteligencia artificial. No quiero dejarles la impresión de que todas las redes neuronales pertenecen a la variedad sencilla de tres capas. Algunos investigadores han continuado estudiando redes de diseños diferentes. En la actualidad, el término *red neuronal* se emplea para

describir un conjunto diverso de modelos, algunos de los cuales son más precisos desde la perspectiva biológica que otros. Pero casi ninguno pretende captar la función o arquitectura general de la corteza cerebral.

En mi opinión, el problema fundamental de la mayor parte de las redes neuronales es un rasgo que comparten con los programas de inteligencia artificial. Ambos soportan la carga fatal de centrar su atención en la conducta. Se denominen dichas conductas "respuestas", "patrones" o "salidas", tanto la inteligencia artificial como las redes neuronales dan por sentado que la inteligencia estriba en la conducta que un programa o una red neuronal produce tras procesar una entrada determinada. El atributo más importante de un programa informático o una red neuronal es si proporciona la salida correcta o deseada. Como por inspiración de Alan Turing, inteligencia es igual a conducta.

Pero la inteligencia no se reduce a actuar o comportarse de modo inteligente. La conducta es una manifestación de inteligencia, pero no la característica central o la definición primordial de ser inteligente. Un momento de reflexión lo demuestra: se puede ser inteligente yaciendo en la oscuridad, pensando y comprendiendo. Ignorar lo que sucede *dentro* de la cabeza y centrarse en la conducta ha constituido un gran impedimento para entender la inteligencia y construir máquinas inteligentes.

\* \* \*

**A**ntes de que exploremos una nueva definición de la inteligencia, quiero hablar de otro planteamiento conectista que estuvo mucho más cerca de describir cómo funcionan los cerebros reales. La pena es que pocas personas parecen haberse dado cuenta de la importancia de esta investigación.

Mientras las redes neuronales acaparaban la atención general, un pequeño grupo escindido de teóricos de dicho campo construía redes que no se centraban en la conducta. Denominadas memorias autoasociativas, también estaban formadas por "neuronas" sencillas que se conectaban entre sí y se estimulaban cuando alcanzaban cierto umbral. Pero estaban interconectadas de modo diferente, utilizando multitud de realimentaciones. En lugar de limitarse a pasar información hacia delante como en una red de propagación recurrente, las memorias autoasociativas alimentaban la salida de cada neurona de nuevo hacia la entrada, algo parecido a llamarte a ti mismo por teléfono. Este bucle de realimentación condujo a algunos rasgos interesantes. Cuando se imponía un patrón de actividad a las neuronas artificiales, formaban una memoria de dicho patrón. La red autoasociativa asociaba patrones consigo misma; de ahí el término *memoria autoasociativa*.

El resultado de este modo de conexión parece ridículo al principio. Para recuperar un patrón almacenado en dicha memoria se debe proporcionar el patrón que se



quiere recuperar. Sería como ir al frutero a comprar un racimo de plátanos. Cuando este te pregunta cómo vas a pagarlo, ofreces hacerlo con plátanos. ¿Qué tiene de bueno?, se preguntarán. Pero una memoria autoasociativa posee algunas propiedades importantes que se encuentran en los cerebros reales.

La propiedad primordial es que no hay que tener el patrón completo que se quiere recuperar para poder hacerlo. Podrías tener solo parte del patrón o un patrón desordenado. La memoria autoasociativa puede recuperar el patrón correcto tal como se almacenó originalmente incluso si aportas una versión desordenada de él. Sería como ir al frutero con unos plátanos marrones a medio comer y obtener a cambio plátanos enteros y verdes. O ir al banco con un billete recortado e ilegible y que el banquero dijera: "Creo que es un billete de 100 dólares estropeado. Démelo y le entregaré este billete nuevo de 100 dólares".

En segundo lugar, a diferencia de la mayor parte de las restantes redes neuronales, se puede diseñar una memoria autoasociativa para que almacene secuencias de patrones o patrones temporales. Este rasgo se logra añadiendo a la realimentación una demora temporal. Con dicha demora se puede presentar a una memoria autoasociativa una secuencia de patrones, similar a una melodía, y es capaz de recordarla. Yo podría aportar a la primera fila unas cuantas notas de *Brilla, brilla, linda estrella*, y la memoria devuelve la canción completa. Cuando se le presenta parte de la secuencia, la memoria es capaz de recordar el resto. Como veremos más adelante, así es como las personas aprenden casi todo, como una secuencia de patrones. Y propongo que el cerebro emplea circuitos similares a una memoria autoasociativa para hacerlo.

Las memorias autoasociativas dieron una idea de la importancia potencial que tenían las entradas con realimentación y cambio de tiempo. Pero la gran mayoría de los científicos cognitivos, de la inteligencia artificial y las redes neuronales pasaron por alto el tiempo y la realimentación.

En su conjunto, los neurocientíficos tampoco lo habían hecho mucho mejor. También conocían la realimentación, pues fueron quienes la descubrieron, pero la mayoría carecía de teoría (más allá de un vago enunciado de fases y modulación) para explicar por qué el cerebro la necesita tanto. Y el tiempo ocupa un papel escaso como mucho en la mayor parte de sus ideas sobre la función general del cerebro. Tienden a representar el cerebro atendiendo a dónde ocurren las cosas, no a cuándo y cómo los patrones neuronales interactúan en el tiempo. Parte de este sesgo proviene de los límites a los que están sometidas nuestras técnicas experimentales actuales. Una de las tecnologías favoritas de la década de 1990, también conocida como la Década del Cerebro, fue la imagen funcional. Las máquinas de imagen funcional pueden tomar fotografías de la actividad cerebral en los humanos, pero no son capaces de observar cambios rápidos. Así pues, los científicos solicitan a los sujetos que se concentren en una sola tarea una y otra vez como si se les pidiera que permanecieran quietos para una fotografía óptica,

con la salvedad de que esta es una fotografía mental. Como resultado contamos con multitud de datos sobre dónde ocurren ciertas tareas en el cerebro, mas pocos sobre cómo fluyen por él entradas reales que varían con el tiempo. La imagen funcional ofrece la oportunidad de ver dónde suceden las cosas en un momento determinado, pero no es capaz de captar fácilmente cómo cambia la actividad cerebral a lo largo del tiempo. Los científicos desearían reunir estos datos, mas existen pocas técnicas buenas para lograrlo. De este modo, muchos neurocientíficos cognitivos de la corriente dominante continúan participando en la falacia de entrada-salida. Se presenta una entrada determinada y se ve qué salida se obtiene. Los diagramas conectados de la corteza cerebral tienden a mostrar mapas de flujos que comienzan en las áreas sensoriales primarias donde entran la vista, los sonidos y el tacto, fluyen por áreas analíticas, planificadoras y motoras superiores, y luego pasan a alimentar de instrucciones a los músculos. Sientes y luego actúas.

No quiero dar a entender que nadie ha tenido en cuenta el tiempo y la realimentación. Se trata de un campo tan enorme que casi cualquier idea cuenta con sus partidarios. En los años recientes ha ido en aumento creer en la importancia de la realimentación, el tiempo y la predicción. Pero el estruendo de la inteligencia artificial y las redes neuronales clásicas mantuvo sometidos e infravalorados los restantes planteamientos durante años.

\* \* \*

**No** es difícil entender por qué la gente —tanto legos como expertos— ha pensado que la conducta define la inteligencia. Durante un par de siglos al menos, se han comparado las capacidades del cerebro con mecanismos de relojería, a continuación con bombas y tuberías, después con motores de vapor y más tarde con ordenadores. Décadas de ciencia ficción han rebosado de ideas sobre la inteligencia artificial, de las leyes de la robótica de Isaac Asimov al C3PO de *La guerra de las galaxias*. La idea de las máquinas que *hacen* cosas está arraigada en nuestra imaginación. Todas las máquinas, estén fabricadas o imaginadas por los humanos, se diseñan para hacer algo. No tenemos máquinas que piensen; tenemos máquinas que hacen. Incluso cuando observamos a nuestros semejantes humanos, nos centramos en su conducta y no en sus pensamientos ocultos. Así pues, parece intuitivamente obvio que la conducta inteligente debe ser la medida de un sistema inteligente.

Sin embargo, si se observa la historia de la ciencia, se comprueba que nuestra intuición suele ser el mayor obstáculo para descubrir la verdad. Los marcos científicos son con frecuencia difíciles de descubrir, no porque sean complejos, sino porque las asunciones intuitivas pero erróneas nos impiden ver la respuesta correcta. Los astrónomos anteriores a Copérnico (1473-1543) erraron al suponer que la Tierra permanecía quieta en el centro del Universo porque *parecía* estar quieta y ocuparlo. Era una intuición evidente que las estrellas formaban todas

parte de una esfera rodante gigantesca con nosotros en el centro. Sugerir que la Tierra giraba como una peonza, moviéndose su superficie a más de 1.600 kilómetros por hora, y que toda ella se desplazaba rápidamente por el espacio — por no mencionar que las estrellas se encuentran a miles de millones de kilómetros de distancia—, te señalaría como un lunático. Pero resultó que ese era el marco acertado. Sencillo de entender, pero erróneo según la intuición.

Antes de Darwin parecía evidente que las especies tenían formas fijas. Los cocodrilos no se emparejan con los colibríes; son distintos e irreconciliables. La idea de que las especies evolucionan no solo iba contra las enseñanzas religiosas, sino también contra el sentido común. La evolución supone que tenemos un antepasado común con cualquier ser vivo de este planeta, incluidos los gusanos y la planta floreciente de nuestra cocina. Ahora sabemos que *tal vez* sea cierto, pero la intuición dice lo contrario.

Menciono estos ejemplos famosos porque creo que la búsqueda de máquinas inteligentes también ha soportado la carga de una asunción intuitiva que está dificultando nuestro progreso. Cuando se pregunta qué hace un sistema inteligente, es evidente que la intuición dicta pensar en la conducta. Demostramos la inteligencia humana mediante el habla, la escritura y las acciones, ¿verdad? Sí; pero solo hasta cierto punto. La inteligencia es algo que sucede en nuestra cabeza. La conducta es un ingrediente opcional. Esto no resulta obvio según la intuición, mas tampoco es difícil de entender.

\* \* \*

**E**n la primavera de 1986, mientras me sentaba ante mi escritorio un día tras otro leyendo artículos científicos, construyendo mi historia de la inteligencia y observando los mundos en evolución de la inteligencia artificial y las redes neuronales, me encontré ahogado en detalles. Había un suministro interminable de cosas que estudiar y leer, pero no estaba logrando ninguna comprensión clara de cómo funciona en realidad el conjunto del cerebro, ni siquiera de qué hacía. Ello se debía a que el mismo campo de la neurociencia estaba inundado de detalles. Y sigue estándolo. Cada año se publican cientos de informes de investigación, pero tienden a añadirse al montón en lugar de organizarlo. Aún no existe una teoría general, un marco, que explique qué hace nuestro cerebro y por qué.

Empecé a imaginarme cómo sería la solución a este problema. ¿Sería extremadamente complicada porque el cerebro es muy complejo? ¿Se precisarían cien páginas de matemática densa para describir cómo funciona el cerebro? ¿Sería necesario representar cientos o miles de circuitos separados antes de que se pudiera comprender algo útil? Pensaba que no. La historia muestra que las mejores soluciones a los problemas científicos son sencillas y elegantes. Aunque puede que los detalles resulten intimidantes y el camino hacia la teoría final sea arduo, por regla general el marco conceptual definitivo es sencillo.

Sin una explicación central que guíe la indagación, los neurocientíficos no cuentan con mucho para proseguir mientras tratan de ensamblar todos los detalles que han reunido para formar un cuadro coherente. El cerebro es increíblemente complejo, una maraña de células vasta y sobrecogedora. A primera vista parece un estadio lleno de espagueti cocido. También se ha descrito como la pesadilla de un electricista. Mas tras una inspección minuciosa vemos que el cerebro no es un montón aleatorio. Posee mucha organización y estructura, pero demasiada para que podamos esperar ser capaces de llegar a intuir el funcionamiento del conjunto de la misma forma que somos capaces de ver cómo los fragmentos de un jarrón roto se vuelven a unir. El fallo no consiste en carecer de datos suficientes o de los datos precisos; lo que necesitamos es un cambio de perspectiva. Con el marco adecuado los detalles cobrarán significado y resultarán manejables. Consideremos la siguiente analogía imaginaria para lograr apreciar lo que quiero decir.

Imaginemos que transcurridos varios milenios los humanos se han extinguido y desembarcan en la Tierra exploradores de una civilización extraterrestre avanzada. Quieren deducir cómo vivíamos. Les intrigan en particular nuestras redes de carreteras. ¿Para qué servían esas estructuras elaboradas y extrañas? Comienzan catalogándolo todo, tanto vía satélite como desde el suelo. Son arqueólogos meticulosos. Registran la situación de cada fragmento perdido de asfalto, cada poste indicador caído y arrastrado monte abajo por la erosión, cada detalle que pueden encontrar. Se dan cuenta de que algunas redes de carreteras son diferentes de otras; en ciertos lugares son sinuosas y estrechas, y su apariencia es casi aleatoria; en otros forman una red regular y en algunos tramos se hacen densas y recorren cientos de kilómetros por el desierto. Recogen una montaña de detalles, pero no significan nada para ellos. Continúan reuniendo más con la esperanza de encontrar algún dato nuevo que lo explique todo. Siguen perplejos durante largo tiempo.

Así están las cosas, hasta que uno de ellos exclama: "¡Eureka! Creo que lo tengo... esas criaturas no podían teletransportarse como nosotros. Tenían que viajar de un lugar a otro, tal vez sobre plataformas móviles con un diseño ingenioso". A partir de esta percepción básica, muchos detalles comienzan a aclararse. Las redes callejeras pequeñas y sinuosas corresponden a las primeras épocas, cuando los medios de transporte eran lentos. Las autopistas densas y largas servían para recorrer grandes distancias a velocidades elevadas y sugerían por fin una explicación de por qué las señales de esas carreteras tenían pintados números diferentes. Los científicos comienzan a deducir las zonas residenciales y las industriales, la forma en que las necesidades del comercio y la infraestructura del transporte debían de haber interactuado, y así sucesivamente. Muchos de los detalles que habían catalogado resultan no ser muy importantes; solo accidentes de la historia o exigencias de la geografía local. Existe la misma cantidad de datos en bruto, pero ya no son desconcertantes.

Cabe confiar en que el mismo tipo de avance nos permitirá comprender qué suponen todos los detalles del cerebro.

\* \* \*

**P**or desgracia, no todos creen que seamos capaces de entender cómo funciona el cerebro. Un sorprendente número de personas, incluidos algunos neurocientíficos, creen que en cierto modo el cerebro y la inteligencia están más allá de toda explicación. Y algunos creen que, aunque lográramos comprenderlos, sería imposible construir máquinas que funcionen del mismo modo, que la inteligencia requiere un cuerpo humano, neuronas, y quizá algunas leyes de la física nuevas e insondables. Siempre que escucho estos argumentos, imagino a los intelectuales del pasado que se negaban al estudio del cielo o se oponían a la disección de los cadáveres para ver cómo funcionaban nuestros cuerpos. "No te molestes en estudiar eso; no llevará a nada bueno y, aunque lograras comprender cómo funciona, no hay nada que se pueda hacer con ese conocimiento." Razonamientos como este nos conducen a una rama de la filosofía llamada funcionalismo, nuestra última parada en esta breve historia de la reflexión sobre el pensamiento.

Según el funcionalismo, ser inteligente o tener mente no es más que una propiedad organizativa, y en esencia carece de importancia qué es lo que confiere la organización. Existe una mente en todo sistema cuyas partes constituyentes poseen la relación causal mutua adecuada, pero esas partes pueden ser con la misma validez neuronas, chips de silicio u otra cosa. Sin duda, esta opinión es la normal para cualquier aspirante a constructor de máquinas inteligentes.

Consideremos lo siguiente: ¿Un juego de ajedrez sería menos real si se jugara con un salero reemplazando a la pieza perdida de un caballo? Es evidente que no. El salero es el equivalente funcional de un caballo "real" en virtud de cómo se mueve sobre el tablero e interactúa con las restantes piezas, así que se trata de un juego de ajedrez verdadero y no de una simulación. O consideremos si esta oración sería la misma si borrara con mi cursor cada uno de los caracteres y volviera a teclearla. O, por tomar un ejemplo más casero, consideremos el hecho de que cada pocos años nuestro cuerpo reemplaza la mayoría de los átomos que lo componen. A pesar de ello, seguimos siendo los mismos en todos los sentidos que nos importan. Un átomo es tan bueno como cualquier otro si desempeña el mismo papel funcional en nuestra constitución molecular. Cabe sostener lo mismo en el caso del cerebro: si un científico loco reemplazara cada una de nuestras neuronas con una réplica de micromáquina funcionalmente equivalente, deberíamos salir del proceso sintiéndonos no menos en nuestro verdadero ser de lo que nos sentíamos al comienzo.

Por este principio, un sistema artificial que emplee la misma arquitectura funcional de un cerebro vivo inteligente debe ser asimismo inteligente y no limitarse a aparentarlo; debe ser real, verdaderamente inteligente.

Los defensores de la inteligencia artificial, los conectistas y yo somos funcionalistas en la medida en que todos creemos que no hay nada inherente, especial o mágico en el cerebro que le permita ser inteligente. Todos creemos que seremos capaces de construir máquinas inteligentes de algún modo cierto día. Pero existen interpretaciones diferentes del funcionalismo. Aunque ya he declarado cuál considero que es el fallo central de la inteligencia artificial y los paradigmas conectistas —la falacia de entrada-salida—, merece la pena decir algo más acerca de por qué todavía no hemos sido capaces de diseñar máquinas inteligentes. Mientras que los defensores de la inteligencia artificial adoptan la que considero una línea dura contraproducente, en mi opinión los conectistas han pecado de tímidos.

Los investigadores de la inteligencia artificial preguntan: “¿Por qué nosotros los ingenieros debemos estar limitados por las soluciones con las que la evolución dio por casualidad?”. En principio, tienen razón. Los sistemas biológicos, como el cerebro y el genoma, se consideran muy poco elegantes. Una metáfora habitual es la de la máquina de Rube Goldberg, bautizada con el nombre del caricaturista de la era de la Depresión que dibujaba artilugios cómicos complicadísimos para realizar tareas nimias. Los diseñadores de *software* cuentan con un término relacionado, *kludge*, para referirse a programas escritos sin previsión que acaban repletos de una complejidad onerosa e inútil, a menudo hasta el punto de volverse incomprensibles incluso para los programadores que los escribieron. Los investigadores de la inteligencia artificial temen que el cerebro sea una confusión similar, un *kludge* de varios millones de años lleno hasta los topes de ineficiencias y un “código heredado” evolucionista. Si es así, se preguntan, ¿por qué no deshacernos de todo ese penoso revoltijo y comenzar de nuevo?

Muchos filósofos y psicólogos cognitivos se muestran favorables a esta postura. Les gusta la metáfora de que la mente se asemeja al *software* que pone en funcionamiento el cerebro, el análogo orgánico del *hardware* informático. En los ordenadores, los niveles de *hardware* y *software* son distintos uno del otro. El mismo programa de *software* funciona en cualquier Máquina de Turing Universal. WordPerfect puede utilizarse en un ordenador personal, un Macintosh o el superordenador Cray, por ejemplo, aunque estos tres sistemas poseen diferentes configuraciones de *hardware*. Y el *hardware* no tiene nada importante que enseñarnos si estamos intentando aprender WordPerfect. Por analogía, prosigue el razonamiento, el cerebro no tiene nada que enseñarnos sobre la mente.

A los defensores de la inteligencia artificial también les gusta señalar ejemplos históricos en los que la solución de la ingeniería difiere radicalmente de la versión de la Naturaleza. Por ejemplo, ¿cómo logramos construir máquinas voladoras?, ¿imitando el movimiento de aleteo de los animales con alas? No. Lo hicimos con alas fijas y árboles de hélice, y más adelante con motores a reacción. Puede que

no sea como lo hizo la Naturaleza, pero funciona, y hasta ahora mejor que batiendo alas.

De modo similar, hicimos un vehículo de tierra que podía correr más que los guepardos, pero no fabricando máquinas de cuatro patas semejantes a dichos animales, sino inventando las ruedas. Son un modo excelente de moverse sobre terreno llano, y el hecho de que la evolución nunca diera con esta estrategia particular no significa que no sea una forma óptima de desplazarnos. Algunos filósofos de la mente se han quedado prendados de la metáfora de la "rueda cognitiva", es decir, la solución de la inteligencia artificial a algún problema que, aunque es totalmente diferente de como lo hace el cerebro, es igual de buena. En otras palabras, un programa que produce salidas que se parecen (o superan) a la ejecución humana de una tarea de modo limitado pero útil es tan bueno como la forma en que nuestros cerebros lo hacen.

Creo que este tipo de interpretación de fines que justifican los medios del funcionalismo descarría a los investigadores de la inteligencia artificial. Como demostró Searle con la habitación china, no basta la equivalencia funcional. Puesto que la inteligencia es una propiedad interna del cerebro, tenemos que mirar dentro de él para entender qué es. En nuestras investigaciones del cerebro, y en especial de la corteza cerebral, necesitaremos ser minuciosos al dilucidar qué detalles son solo "accidentes congelados" superfluos de nuestro pasado evolutivo; sin duda, muchos procesos del tipo de Rube Goldberg están mezclados con los rasgos importantes. Pero, como veremos enseguida, existe una elegancia subyacente de gran potencia, una que supera a nuestros mejores ordenadores, esperando a ser extraída de esos circuitos neuronales.

Los conectistas percibieron intuitivamente que el cerebro no era un ordenador y que sus secretos radicaban en el modo de comportarse sus neuronas cuando se conectaban entre sí. Fue un buen comienzo, pero el campo apenas ha avanzado desde sus primeros logros. Aunque miles de personas trabajaron en redes de tres capas, y muchas siguen haciéndolo, la investigación de redes corticales realistas fue y sigue siendo rara.

Durante medio siglo hemos estado aplicando toda la fuerza del considerable ingenio de nuestra especie a tratar de programar inteligencia en los ordenadores. En el proceso se nos han ocurrido procesadores de texto, bases de datos, videojuegos, Internet, teléfonos móviles y convincentes dinosaurios animados por ordenador. Pero las máquinas inteligentes siguen sin aparecer en el cuadro. Para tener éxito necesitaremos copiar mucho del motor de la inteligencia de la Naturaleza, la corteza cerebral. Tenemos que extraer la inteligencia del cerebro. Ningún otro camino nos conducirá ahí.

### 3

#### El cerebro Humano

**Así** pues, ¿qué hace al cerebro tan diferente de la programación que se incorpora en la inteligencia artificial y las redes neuronales? ¿Qué tiene de inusual el diseño del cerebro y por qué es importante? Como veremos en los próximos capítulos, la arquitectura cerebral tiene mucho que decirnos sobre cómo funciona el cerebro y por qué es radicalmente diferente de un ordenador.

Comencemos nuestra introducción con el órgano en su conjunto. Imaginemos que hay un cerebro colocado sobre una mesa y que lo disecamos juntos. Lo primero que apreciamos es que su superficie exterior parece muy uniforme. De un gris rosado, se asemeja a una coliflor suave, con numerosas crestas y valles, llamados circunvoluciones y surcos. Es blando y húmedo al tacto. Se trata de la corteza cerebral, una delgada capa de tejido neuronal que envuelve la mayoría de las partes más antiguas del cerebro. En ella vamos a centrar nuestra atención particular. Casi todo lo que pensamos que es inteligencia —la percepción, el lenguaje, la imaginación, la matemática, el arte, la música y la planificación— ocurre ahí. Su corteza cerebral está leyendo este libro.

Ahora tengo que admitir que soy un chovinista de la corteza cerebral. Sé que voy a encontrar cierta resistencia por ello, así que permítanme un minuto para defender mi postura antes de proseguir explicándola. Cada parte del cerebro posee su propia comunidad de científicos que la estudian, y la sugerencia de que podamos llegar al fondo de la inteligencia entendiendo solo la corteza cerebral sin duda suscitará unos cuantos alaridos y objeciones de las comunidades de investigadores ofendidos. Dirán cosas como: “No es posible entender la corteza cerebral sin comprender la región cerebral *tal*, porque las dos están muy interconectadas, y necesita la región cerebral *tal* para hacer esto y lo otro”. Estoy de acuerdo. Concedo que el cerebro consta de muchas partes y la mayoría son cruciales para el ser humano. (Una curiosa excepción es la parte del cerebro con el mayor número de células, el cerebelo. Si se nace sin cerebelo o está dañado, se puede llevar una vida normal. Sin embargo, no ocurre lo mismo con la mayoría de las regiones restantes, que se requieren para la vida básica o el estado consciente.)

Mi razonamiento es que no estoy interesado en construir humanos. Quiero entender la inteligencia y construir máquinas inteligentes. Ser humano y ser inteligente son asuntos separados. Una máquina inteligente no necesita tener impulsos sexuales, hambre, pulso, músculos, emociones o cuerpo semejante al humano. Un humano es mucho más que una máquina inteligente. Somos criaturas biológicas con todo lo necesario y a veces el bagaje indeseado que proviene de eones de evolución. Si se desea construir máquinas inteligentes que se comporten como humanos —es decir, que pasen el test de Turing en todos sus aspectos— es probable que se tenga que recrear buena parte de la restante composición que



hace a los humanos como son. Pero, como veremos más adelante, para construir máquinas que sean inteligentes de verdad, pero no exactas a los humanos, podemos centrarnos en la parte del cerebro estrictamente relacionada con la inteligencia.

A aquellos que tal vez se sientan ofendidos por la atención singular que presto a la corteza cerebral, les diré que estoy de acuerdo en que otras estructuras cerebrales, como el tronco encefálico, los ganglios basales y el núcleo amigdalino, son importantes para el funcionamiento de aquella. Sin duda alguna. Pero espero convencerlos de que todos los aspectos esenciales de la inteligencia ocurren en la corteza cerebral, aunque también desempeñen importantes papeles otras regiones cerebrales, el tálamo y el hipotálamo, de los que nos ocuparemos más adelante en el libro. A largo plazo, necesitaremos comprender los papeles funcionales de todas las regiones cerebrales. Pero creo que esos temas se abordarán mejor en el contexto de una buena teoría general de la función de la corteza cerebral. Esta es mi opinión resumida, y ahora volvamos a la corteza cerebral.

Cojan seis tarjetas de visita o cartas de la baraja —cualesquiera de ellas valdrán— y pónganlas en un montón. (Convendría que lo hicieran de verdad en lugar de limitarse a imaginarlo.) Ahora cuentan con un modelo de la corteza cerebral. Las seis tarjetas tienen un espesor de unos dos milímetros y les proporcionarán el sentido de lo delgada que es la lámina cortical. Al igual que el montón de tarjetas o cartas, la corteza cerebral tiene un grosor aproximado de dos milímetros y cuenta con seis capas, más o menos una por cada tarjeta o carta.

Extendida, la lámina de la corteza cerebral humana alcanza el tamaño aproximado de una servilleta grande. Las láminas corticales de otros mamíferos son menores: la de la rata es del tamaño de un sello de correos; la del mono, del tamaño de un sobre de una carta comercial. Pero, dejando a un lado el tamaño, la mayoría de ellas contienen seis capas similares a las que vemos en el montón de tarjetas de visita. Los humanos somos más listos porque nuestra corteza cerebral, en relación con el tamaño corporal, ocupa una zona mayor, no porque nuestras capas sean más gruesas o contengan alguna clase especial de células "listas". Su tamaño es bastante impresionante, pues rodea y envuelve la mayor parte del resto del cerebro. Para acomodar nuestro gran cerebro, la Naturaleza tuvo que modificar nuestra anatomía general. Las hembras humanas desarrollaron una pelvis ancha para dar a luz niños de cabeza grande, rasgo que algunos paleoantropólogos piensan que coevolucionó con la capacidad de caminar sobre las dos piernas. Pero como seguía sin ser suficiente, la evolución plegó la corteza cerebral, metiéndola dentro de nuestros cráneos como una hoja de papel arrugada dentro de una copa de coñac.

La corteza cerebral está cargada de células nerviosas o neuronas. Están tan apiñadas que nadie sabe con precisión cuántas células contiene. Si dibujan un cuadrado diminuto de un milímetro de lado (aproximadamente la mitad del tamaño

de esta letra o) en la parte superior del montón de tarjetas de visita, están marcando la posición estimada de cien mil neuronas. Imaginen tratar de contar el número exacto en un espacio tan reducido; es casi imposible. No obstante, algunos anatomistas han calculado que la corteza cerebral humana media contiene alrededor de treinta mil millones de neuronas, pero nadie se sorprendería si la cifra fuera mucho más alta o baja.

Esos treinta mil millones de células es usted. Contienen casi todos sus recuerdos, conocimientos, capacidades y experiencia vital acumulada. Tras veinticinco años de pensar en los cerebros, este hecho me sigue pareciendo asombroso. Que una delgada lámina de células vea, sienta y cree nuestra visión del mundo es algo increíble. El calor de un día veraniego y los sueños que tenemos de un mundo mejor son en cierto modo la creación de estas células. Muchos años después de haber publicado su artículo en *Scientific American*, Francis Crick escribió un libro sobre cerebros llamado *The Astonishing Hypothesis*. La hipótesis asombrosa era simplemente que la mente es la creación de las células del cerebro. No hay nada más, nada mágico, ninguna salsa especial; solo neuronas y un baile de información. Espero que sean capaces de percibir lo increíble que resulta darse cuenta de ello. Parece haber un gran abismo filosófico entre una reunión de células y nuestra experiencia consciente, si bien la mente y el cerebro son la misma cosa. Al denominarlo una hipótesis, Crick se mostraba políticamente correcto. Que las células de nuestro cerebro crean la mente es un hecho, no una hipótesis. Necesitamos comprender qué hacen esos treinta mil millones de células y cómo lo hacen. Por suerte, la corteza cerebral no es solo una gota de células amorfa. Podemos observarla en profundidad para buscar ideas sobre cómo se comporta la mente humana.

\* \* \*

**V**olvamos a nuestra mesa de disección y miremos un poco más el cerebro. A simple vista, la corteza cerebral no ofrece casi señales. Sin duda, hay algunas, como la enorme fisura que separa los dos hemisferios cerebrales y el surco prominente que divide las regiones posterior y frontal. Pero se mire donde se mire, de la izquierda a la derecha y de atrás adelante, la superficie enrollada resulta muy parecida. No existen líneas limítrofes visibles o códigos de color que delimiten zonas especializadas en información sensorial diferente o distintos tipos de pensamiento.

Sin embargo, la gente ha sabido desde hace tiempo que existe algún tipo de frontera. Incluso antes de que los neurocientíficos fueran capaces de discernir algo útil sobre el sistema de circuitos de la corteza cerebral, sabían que algunas funciones mentales estaban localizadas en ciertas regiones. Si una apoplejía deja fuera de combate el lóbulo parietal derecho de Joe, puede perder su capacidad de percibir —o incluso de concebir— cualquier cosa del lado izquierdo de su cuerpo, o de la mitad izquierda del espacio a su alrededor. En contraste, una apoplejía en la

región frontal izquierda, conocida como área de Broca, compromete su capacidad de emplear las reglas gramaticales, aunque su vocabulario y su facultad para entender los significados de las palabras no cambien. Una apoplejía en una zona llamada circunvolución fusiforme puede acabar con la capacidad de reconocer rostros: Joe no puede reconocer a su madre, a sus hijos y ni siquiera su propia cara en una fotografía. Trastornos tan fascinantes como estos facilitaron pronto a los neurocientíficos la noción de que la corteza cerebral consta de muchas regiones o áreas funcionales. Los términos son equivalentes.

Hemos aprendido bastante sobre áreas funcionales en el siglo pasado, pero queda mucho por descubrir. Cada una de estas regiones es semiindependiente y parece estar especializada para ciertos aspectos de la percepción o el pensamiento. Físicamente, están dispuestas como una serie de parches irregulares que varía poco de una persona a otra. Rara vez las funciones están delimitadas con claridad. Desde la perspectiva funcional, están organizadas en una jerarquía con ramificaciones.

La noción de jerarquía es crucial, así que quiero dedicar un momento a definirla con cuidado, pues me estaré refiriendo a ella a lo largo de todo el libro. En un sistema jerárquico, algunos elementos están en un sentido abstracto "encima" y "debajo" de otros. En una jerarquía empresarial, por ejemplo, un gerente de nivel medio está por encima del empleado que se encarga del correo y por debajo del vicepresidente. Esto no tiene nada que ver con la altura o bajura física; aunque trabaje en un piso inferior al del encargado del correo, el gerente sigue estando por "encima" desde el punto de vista jerárquico. Recalco este punto para aclarar lo que quiero decir siempre que hablo de que una región funcional está más alta o más baja que otra. No tiene nada que ver con su disposición física en el cerebro. Todas las áreas funcionales de la corteza cerebral residen en la misma lámina cortical. Lo que hace que una región sea "superior" o "inferior" a otra es su forma de conectarse entre sí. En la corteza cerebral, las áreas inferiores alimentan de información a las superiores mediante un patrón neuronal de conectividad, mientras que las áreas superiores envían realimentación "hacia abajo" a las áreas inferiores empleando un patrón de conexión diferente. También existen conexiones laterales entre áreas que están en ramas separadas de la jerarquía, del mismo modo que un gerente de nivel medio se comunica con su semejante en un despacho asociado de otro estado. Dos científicos, Daniel Felleman y David van Essen, han elaborado un mapa detallado de la corteza cerebral del mono. Dicho mapa muestra docenas de regiones conectadas entre sí en una jerarquía compleja. Cabe asumir que la corteza cerebral humana posee una jerarquía similar.

Las regiones funcionales más bajas, las áreas sensoriales primarias, es el lugar donde llega primero la información sensorial. Procesan la información en su nivel más bruto y básico. Por ejemplo, la información visual entra en la corteza cerebral a través del área visual primaria, llamada V1 para abreviar. V1 está conectada con

rasgos visuales de nivel inferior, como segmentos de bordes diminutos, componentes de pequeña escala del movimiento, disparidad binocular (para la visión estéreo) e información básica de color y contraste. V1 suministra información a las áreas V2, V4 e IT (hablaremos sobre ellas más adelante), y a muchas otras regiones. Cada una de dichas áreas se ocupa de aspectos más especializados o abstractos de la información. Por ejemplo, las células de V4 responden a objetos de complejidad media, como formas de estrella en diferentes colores como rojo o azul. Otra área llamada MT se especializa en los movimientos de los objetos. En los escalones más elevados de la corteza cerebral visual se encuentran áreas que representan recuerdos visuales de todo tipo de objetos, como rostros, animales, herramientas, partes del cuerpo y demás.

Los restantes sentidos presentan jerarquías similares. La corteza cerebral posee un área auditiva primaria llamada A1 y una jerarquía de regiones auditivas por encima, y cuenta con un área somatosensorial primaria (sentido corporal) llamada S1 y una jerarquía de regiones somatosensoriales por encima. Al final la información sensorial pasa a "áreas de asociación", que es el nombre que a veces se emplea para las regiones de la corteza cerebral que reciben entradas de más de un sentido. Por ejemplo, nuestra corteza cerebral tiene áreas que reciben entradas tanto de la visión como del tacto. Gracias a las regiones de asociación somos capaces de darnos cuenta de que la visión de una mosca andando por nuestro brazo y la sensación de cosquillas que sentimos comparten la misma causa. La mayoría de estas áreas reciben entradas muy procesadas de varios sentidos, y sus funciones continúan sin estar claras. Más adelante en el libro tendré mucho que decir sobre la jerarquía cortical.

Existe un conjunto más de áreas en los lóbulos frontales del cerebro que crean salidas motoras. El sistema motor de la corteza cerebral también está organizado según una jerarquía. El área inferior, M1, envía conexiones a la médula espinal y maneja los músculos de forma directa. Las áreas superiores alimentan de órdenes motoras complejas a M1. La jerarquía del área motora y las jerarquías de las áreas sensoriales resultan muy similares. Parecen estar armadas de la misma forma. En la región motora pensamos en la información que fluye hacia abajo de la jerarquía hasta M1 para manejar los músculos, y en las regiones sensoriales pensamos en la información que fluye hacia arriba de la jerarquía alejándose de los sentidos. Pero en realidad la información fluye en ambas direcciones. Lo que en las regiones sensoriales se advierte como realimentación es la salida de la región motora, y viceversa.

La mayoría de las descripciones de los cerebros se basan en mapas de flujos que reflejan una visión de las jerarquías muy simplificada. Esto es, la entrada (visiones, sonidos, tactos) fluye a las áreas sensoriales primarias y se procesa mientras avanza hacia arriba de la jerarquía, luego pasa por las áreas de asociación, a continuación por los lóbulos frontales de la corteza cerebral, y por último

desciende a las áreas motoras. No afirmo que esta visión sea completamente equivocada. Cuando leemos en voz alta, la información visual entra en V1, fluye hasta las áreas de asociación, se abre paso hasta la corteza motora frontal y concluye haciendo que los músculos de nuestra boca y garganta formen los sonidos del habla. Sin embargo, esto no es todo lo que ocurre. No es tan sencillo. En la visión demasiado simplificada contra la que prevengo, el proceso se suele tratar como si la información fluyera en una sola dirección, como los adminículos que se arman en la cadena de montaje de una fábrica. Pero la información en la corteza cerebral siempre fluye también en la dirección contraria, y con muchas más proyecciones alimentando hacia abajo de la jerarquía que hacia arriba. Cuando leemos en voz alta, las regiones superiores de nuestra corteza cerebral envían más señales hacia "abajo" a nuestra corteza visual primaria que las que nuestro ojo recibe de la página impresa. Nos ocuparemos en capítulos posteriores de lo que esas proyecciones de realimentación hacen. De momento, quiero que graben un hecho: aunque la jerarquía ascendente es real, tenemos que tener cuidado para no pensar que el flujo de información solo tiene una dirección.

De nuevo en la mesa de disección, supongamos que colocamos un potente microscopio, cortamos una rebanada fina de la lámina cortical, teñimos algunas células y echamos una ojeada a nuestra obra a través del ocular. Si teñimos todas las células de nuestra rebanada, veremos una masa toda negra porque las células están muy apiñadas y entrelazadas. Pero si empleamos un tinte que marque una fracción más pequeña de células, podemos ver las seis capas que he mencionado. Estas capas están formadas por variaciones de la densidad de las células corporales, los tipos de células y sus conexiones.

Todas las neuronas poseen rasgos en común. Además del cuerpo celular, que es la parte redondeada que imaginamos cuando pensamos en una célula, también tienen estructuras ramificadas, semejantes a alambres, llamadas axones y dendritas. Cuando el axón de una neurona toca la dendrita de otra, se forman pequeñas conexiones llamadas sinapsis. En las sinapsis es donde el impulso nervioso de una célula influye la conducta de otra célula. Si llega una punta a una sinapsis, es posible que también llegue a la célula receptora. Algunas sinapsis tienen el efecto contrario y hacen menos probable que la célula receptora también genere una punta. La fuerza de una sinapsis puede cambiar según la conducta de las dos células. La forma más simple de este cambio sináptico es el aumento de la fuerza de conexión entre dos neuronas cuando ambas generan una punta casi al mismo tiempo. Analizaré más este proceso, denominado aprendizaje hebbiano, un poco más adelante. Además de cambiar la fuerza de una sinapsis, hay pruebas que indican que pueden formarse sinapsis completamente nuevas entre dos neuronas. Tal vez suceda de forma continua, aunque las pruebas científicas son polémicas. Dejando a un lado los detalles sobre cómo cambian sus fuerzas las sinapsis, lo que sí es seguro es que la formación y la fuerza de las sinapsis es lo que hace que las memorias se almacenen.

Aunque hay muchos tipos de neuronas en la corteza cerebral, una amplia clase comprende ocho de cada diez células. Se trata de las neuronas piramidales, así llamadas porque sus cuerpos celulares presentan una forma parecida a las pirámides. Salvo la capa superior de las seis que forman la corteza cerebral, que tiene miles de axones pero muy pocas células, las restantes contienen células piramidales. Cada neurona piramidal se conecta con muchas otras de sus inmediaciones, y cada una envía un largo axón lateral a regiones más distantes de la corteza cerebral o a estructuras cerebrales inferiores como el tálamo.

Una célula piramidal común posee varios miles de sinapsis. Una vez más, resulta muy difícil saber con exactitud cuántas debido a su extrema densidad y reducido tamaño. El número de sinapsis varía de una célula a otra, de una capa a otra y de una región a otra. Si aceptamos la posición conservadora de que la célula piramidal media tiene mil sinapsis (el número real es probable que se aproxime a las cinco o diez mil), nuestra corteza cerebral se acercaría a los treinta mil billones de sinapsis. Es un número astronómico que sobrepasa con creces nuestra comprensión intuitiva. Parece bastar para guardar todas las cosas que se pueden aprender en una vida.

\* \* \*

**Según** el rumor, Albert Einstein afirmó una vez que concebir la teoría de la relatividad especial había sido algo inmediato, casi fácil. La dedujo de forma natural de una sola observación: la velocidad de la luz es constante para todos los observadores, aun cuando estos se muevan a velocidades diferentes. Lo cual va contra la intuición. Es como decir que la velocidad de una pelota lanzada es siempre la misma prescindiendo de la fuerza con la que se lance o de la rapidez con la que se muevan los individuos que la lanzan y observan. Todos ven la pelota moviéndose a la misma velocidad en relación con ellos en todas las circunstancias. No parece que pueda ser cierto, pero se demostró que así era por la luz; y el inteligente Einstein se preguntó cuáles eran las consecuencias de este extraño hecho. Reflexionó metódicamente sobre todas las repercusiones de una velocidad de la luz constante, lo que le condujo a las predicciones aún más extrañas de la relatividad especial, tales como que el tiempo se enlentecía cuando se avanzaba más deprisa, y que la energía y la masa eran en esencia lo mismo. Los libros sobre la relatividad recorren esta línea de razonamiento con ejemplos cotidianos de trenes, balas, linternas y demás. La teoría no es difícil, pero sin duda va contra la intuición.

Existe un descubrimiento análogo en la neurociencia, un hecho sobre la corteza cerebral que resulta tan sorprendente que algunos neurocientíficos se niegan a creerlo y la mayoría restante lo pasa por alto porque no sabe qué hacer con él. Pero es un hecho de tal importancia que si se exploran sus consecuencias cuidadosa y metódicamente desvelará los secretos de lo que hace la corteza cerebral y cómo funciona. En este caso, el descubrimiento sorprendente provino de

la anatomía básica de la propia corteza, pero fue precisa una mente con una perspicacia fuera de lo común para reconocerlo. Esa persona fue Vernon Mountcastle, neurocientífico de la Universidad John Hopkins de Baltimore. En 1978 publicó un artículo titulado "An Organizing Principle for Cerebral Function", en el que señala que la corteza cerebral es notablemente uniforme en cuanto a apariencia y estructura. Las regiones que se ocupan de las entradas auditivas se asemejan a las que se ocupan del tacto, que se parecen a las regiones que controlan los músculos, semejantes al área de lenguaje de Broca, que es parecida a casi todas las restantes regiones de la corteza cerebral. Mountcastle sugiere que puesto que dichas regiones parecen semejantes, tal vez realicen la misma operación básica. Propone que la corteza cerebral usa la misma herramienta computacional para realizar todo lo que hace.

Los anatomistas de la época y de las décadas anteriores a Mountcastle reconocían que la corteza cerebral parece toda semejante; es algo innegable. Pero en lugar de preguntarse qué podía significar, dedicaron su tiempo a buscar diferencias entre un área y otra. Y las encontraron. Supusieron que si una región se emplea para el lenguaje y otra para la visión, debe haber diferencias entre ambas. Si se buscan con el cuidado suficiente, se encuentran. Las regiones de la corteza cerebral varían en espesor, densidad celular, proporción relativa de células y muchos otros aspectos que pueden resultar difíciles de descubrir. Una de las regiones más estudiadas, el área visual primaria V1, presenta unas cuantas divisiones más en una de sus capas. La situación es análoga a la labor de los biólogos a mediados de la década de 1800. Dedicaron su tiempo a descubrir las diferencias mínimas entre las especies. El éxito consistía en descubrir que dos ratones que parecían casi idénticos eran en realidad especies separadas. Durante muchos años, Darwin siguió el mismo curso, estudiando con frecuencia moluscos. Pero acabó teniendo la gran percepción de preguntarse cómo todas esas especies podían ser tan parecidas. Es su similitud la que resulta sorprendente e interesante, mucho más que sus diferencias.

Mountcastle realiza una observación similar. En un campo de anatomistas que buscan diferencias mínimas en las regiones corticales, muestra que, a pesar de las diferencias, la corteza cerebral es notablemente uniforme. Las mismas capas, tipos de células y conexiones existen por todas partes. Se parece toda ella a las seis tarjetas de visita. Las diferencias son a menudo tan sutiles que los anatomistas entrenados no pueden ponerse de acuerdo al respecto. Por lo tanto, sostiene Mountcastle, todas las regiones de la corteza cerebral ejecutan las mismas operaciones. Lo que hace que el área de la visión sea visual y el área del movimiento sea motora es el modo como se conectan entre sí y con otras partes del sistema nervioso central las diversas regiones de la corteza cerebral.

De hecho, Mountcastle sostiene que la razón de que una región de la corteza parezca ligeramente distinta de otra son sus conexiones, y no se debe a que su

función básica sea diferente. Concluye que existe una función común, un algoritmo común que ejecutan todas las regiones corticales. La visión no es diferente del oído, que no es diferente de una salida motora. Concede que nuestros genes especifican cómo están conectadas las regiones, lo cual es muy peculiar de la función y la especie, pero el tejido cortical en sí hace lo mismo en todos los lugares.

Reflexionemos sobre ello un momento. Para mí, la vista, el oído y el tacto parecen muy diferentes. Poseen cualidades esencialmente distintas. La vista supone color, textura, contorno, profundidad y forma. El oído tiene tono, ritmo y timbre. Parecen muy diferentes. ¿Cómo pueden ser lo mismo? Mountcastle afirma que no lo son, pero el modo como la corteza cerebral procesa las señales procedentes del oído es el mismo que emplea para procesar las señales de los ojos. Prosigue diciendo que el control motor funciona también según el mismo principio.

En su mayor parte, los científicos e ingenieros han ignorado la propuesta de Mountcastle, o han preferido pasarla por alto. Cuando tratan de entender la visión o fabricar un ordenador capaz de "ver", idean vocabulario y técnicas específicas para la visión. Hablan de bordes, texturas y representaciones tridimensionales. Si quieren comprender el lenguaje hablado, construyen algoritmos basados en reglas gramaticales, sintaxis y semántica. Pero si Mountcastle está en lo cierto, dichos planteamientos no se ajustan al modo como el cerebro resuelve estos problemas, y por lo tanto es probable que fracasen. Si Mountcastle tiene razón, el algoritmo de la corteza cerebral debe expresarse con independencia de toda función o sentido particular. El cerebro emplea el mismo proceso para ver que para oír. La corteza cerebral hace algo universal que puede aplicarse a cualquier tipo de sistema sensorial o motor.

Cuando leí por primera vez el artículo de Mountcastle, casi me caigo de la silla. Ahí estaba la piedra Rosetta de la neurociencia, un único artículo y una única idea que unían todas las facultades diversas y maravillosas de la mente humana. Las unía bajo un algoritmo único. Con un solo paso dejaba al descubierto la falacia de todos los intentos previos de entender y urdir la conducta humana como capacidades diversas. Espero que sean capaces de apreciar la elegancia radical y maravillosa de la propuesta de Mountcastle. Las mejores ideas de la ciencia siempre son sencillas, elegantes e inesperadas, y esta es una de las mejores. En mi opinión, fue, es y probablemente seguirá siendo el descubrimiento más importante de la neurociencia. Sin embargo, por increíble que resulte, la mayoría de los científicos e ingenieros se niegan a creer en ella, prefieren pasarla por alto o no la conocen.

\* \* \*

**P**arte de esta desidia se debe a la escasez de herramientas para estudiar cómo fluye la información dentro de las seis capas de la corteza cerebral. Las herramientas con que contamos operan a un nivel burdo y en general su objetivo



consiste en determinar dónde —frente a cuándo y cómo— surgen las diversas facultades en la corteza cerebral. Por ejemplo, buena parte de la neurociencia que aparece en la prensa popular de nuestros días favorece de forma implícita la idea de que el cerebro es una reunión de módulos de alta especialización. Las técnicas de imagen funcional como los escáneres funcionales MRI y PET se centran casi con exclusividad en los mapas cerebrales y las regiones funcionales que ya he mencionado. En estos experimentos, un sujeto voluntario está tumbado con la cabeza dentro del escáner y ejecuta una clase de tarea mental o motora. Podría estar jugando a un videojuego, generando conjugaciones verbales, leyendo oraciones, mirando rostros, describiendo fotos, imaginando algo, memorizando listas, tomando decisiones financieras y demás. El escáner detecta qué regiones del cerebro están más activas de lo habitual durante estas tareas y dibuja manchas coloreadas sobre una imagen del cerebro del sujeto para indicarlas. Al parecer, estas regiones son centrales para la tarea. Se han realizado miles de experimentos de imágenes funcionales, y seguirán miles más. Con todos ellos estamos construyendo poco a poco un cuadro indicador de dónde suceden ciertas funciones en el cerebro adulto normal. Es fácil afirmar: “esta es el área de reconocimiento de caras, esta es el área de la matemática, esta es el área de la música”, y así sucesivamente. Como no sabemos cómo ejecuta el cerebro dichas tareas, resulta natural asumir que lo hace de modos diferentes.

¿Pero es así? Un conjunto creciente y fascinante de pruebas apoya la propuesta de Mountcastle. Algunos de los mejores ejemplos demuestran la extrema flexibilidad y plasticidad de la corteza cerebral. Todo cerebro humano, si se nutre adecuadamente y se coloca en el entorno preciso, puede aprender una cualquiera de las miles de lenguas habladas. Ese mismo cerebro es capaz también de aprender el lenguaje de los signos, el lenguaje escrito, el lenguaje musical, el lenguaje matemático, los lenguajes informáticos y el lenguaje corporal. Puede aprender a vivir en los helados climas del Norte o en un desierto abrasador. Puede llegar a ser un experto en ajedrez, pesca, agricultura o física teórica. Consideremos el hecho de que tengamos una pequeña área visual especial que parece dedicada a representar letras y dígitos escritos. ¿Significa que hemos nacido con un área de lenguaje dispuesta para procesar letras y dígitos? No es muy probable. El lenguaje escrito es un invento demasiado reciente para que nuestros genes hayan evolucionado un mecanismo específico al respecto. Así pues, la corteza cerebral todavía sigue dividiéndose en áreas funcionales con tareas específicas hasta bien avanzada la infancia, basadas puramente en la experiencia. El cerebro humano posee una capacidad increíble de aprender y adaptarse a miles de entornos que no existían hasta fecha muy reciente. Lo cual aboga por un sistema con una flexibilidad increíble, no por uno con miles de soluciones para miles de problemas.

Los neurocientíficos también han descubierto que el sistema de conexiones de la corteza cerebral es sorprendentemente “plástico”, lo que significa que puede cambiar y reconectarse según el tipo de entradas que reciba. Por ejemplo, a los

hurones recién nacidos se les pueden cambiar quirúrgicamente las conexiones para que los ojos del animal envíen sus señales a las áreas de la corteza cerebral donde normalmente se desarrolla el oído. El resultado sorprendente es que desarrollan sendas visuales funcionales en las porciones auditivas de sus cerebros. En otras palabras, ven con tejido cerebral que normalmente escucha sonidos. Se han efectuado experimentos similares con otros sentidos y regiones cerebrales. Por ejemplo, al poco de nacer, se pueden trasplantar trozos de corteza cerebral visual de rata a regiones en las que se suele representar el sentido del tacto. Cuando la rata crece, el tejido trasplantado procesa el tacto en lugar de la visión. Las células no han nacido para especializarse en visión, tacto o audición.

La corteza cerebral humana es igual de plástica. Los adultos que nacen sordos procesan la información visual en áreas que normalmente se vuelven regiones auditivas. Y los adultos ciegos de nacimiento usan la parte de más atrás de su corteza cerebral, que por lo general se dedica a la visión, a leer braille. Como el braille supone tacto, cabría pensar que en principio activaría las regiones del tacto, pero al parecer ningún área de la corteza cerebral se contenta con representar nada. La corteza cerebral visual, al no recibir información de los ojos como se "supone", trata de encontrar alrededor otros patrones de entrada para pasarse a ellos; en este caso, de otras regiones corticales.

Todo esto pretende mostrar que las regiones cerebrales desarrollan funciones especializadas basadas en buena medida en el tipo de información que fluye a ellas durante el desarrollo. La corteza cerebral no está diseñada de forma rígida para realizar diferentes funciones utilizando algoritmos distintos, del mismo modo que la superficie de la Tierra no estaba predestinada a acabar con su ordenamiento moderno de naciones. La organización de nuestra corteza cerebral, al igual que la geografía política del globo, podría haber resultado diferente si se hubieran dado un conjunto de circunstancias distintas.

Los genes dictan la arquitectura general de la corteza cerebral, incluidas las especificaciones de qué regiones se conectan entre sí, pero dentro de esa estructura el sistema es muy flexible.

Mountcastle estaba en lo cierto. No hay más que un único algoritmo puesto en práctica por cada una de las regiones del cerebro. Si se conectan las regiones de la corteza cerebral con una jerarquía apropiada y se proporciona un flujo de entrada, aprenderán de su entorno. Por lo tanto, no hay razón para que las máquinas inteligentes del futuro tengan los mismos sentidos o facultades que nosotros los humanos. El algoritmo cortical puede utilizarse de modos novedosos, con sentidos nuevos, en una lámina cortical fabricada de manera que surja una inteligencia verdadera y flexible fuera de los cerebros biológicos.

\* \* \*

**P**asemos ahora a un tema que está relacionado con la propuesta de Mountcastle y es igual de sorprendente. Las entradas que recibe nuestra corteza cerebral son todas básicamente iguales. Una vez más, es probable que piensen que sus sentidos son entidades bien separadas. Después de todo, el sonido se transporta como ondas de compresión por el aire; la visión, como luz, y el tacto, como una presión sobre la piel. El sonido parece temporal; la visión, sobre todo pictorial; y el tacto, espacial. ¿Qué podría ser más diferente que el sonido de una cabra balando frente a la visión de una manzana y frente al tacto de una pelota de béisbol?

Pero observemos con mayor detenimiento. La información visual procedente del mundo exterior se envía a nuestro cerebro vía un millón de fibras del nervio óptico. Después de un breve tránsito por el tálamo, llega a la corteza cerebral visual primaria. Los sonidos se envían vía las treinta mil fibras del nervio auditivo. Pasan por algunas partes más antiguas del cerebro y luego llegan a la corteza cerebral auditiva primaria. La médula espinal traslada información del tacto y las sensaciones internas al cerebro vía otro millón de fibras, que son recibidas por la corteza cerebral somatosensorial primaria. Estas son las principales entradas de nuestro cerebro. Son como sentimos el mundo.

Cabe visualizar estas entradas como un lío de cables eléctricos o un manojo de fibras ópticas. Tal vez hayan visto lámparas hechas con fibras ópticas donde aparecen puntos de luz coloreada al final de cada una. Las entradas al cerebro son semejantes, pero las fibras se denominan axones y transportan señales neuronales llamadas "potenciales de acción" o "puntas", que son en parte químicas y en parte eléctricas. Los órganos sensoriales que suministran dichas señales son diferentes, pero una vez que se han convertido en potenciales de acción dirigidos al cerebro son todos lo mismo: nada más que patrones.

Si miramos un perro, por ejemplo, un conjunto de patrones fluirá por las fibras de nuestro nervio óptico hasta la parte visual de la corteza cerebral. Si escuchamos ladrar al perro, fluirá un conjunto diferente de patrones por nuestro nervio auditivo hasta las partes auditivas del cerebro. Si acariciamos al perro, un conjunto de patrones de tacto-sensación fluirá de nuestra mano a través de las fibras de la médula espinal hasta las partes del cerebro que se ocupan del tacto. Cada patrón —ver el perro, escuchar al perro, sentir al perro— se experimenta de forma diferente porque cada uno se canaliza por una senda diferente en la jerarquía cortical. Tiene importancia dónde entran en el cerebro los cables. Pero en el nivel abstracto de las entradas sensoriales son todos en esencia lo mismo, y todos se manejan de forma similar por las seis capas de la corteza cerebral. Escuchamos el sonido, vemos la voz y sentimos la presión, mas dentro de nuestro cerebro no existe ninguna diferencia fundamental entre esos tipos de información. Una acción potencial es una acción potencial. Estas puntas momentáneas son idénticas, prescindiendo de lo que originariamente las causó. Todo nuestro cerebro conoce estos patrones.

Nuestras percepciones y conocimiento sobre el mundo se construyen con estos patrones. No hay luz dentro de nuestras cabezas; hay oscuridad. Tampoco entra sonido en el cerebro; dentro hay silencio. De hecho, el cerebro es la única parte de nuestro cuerpo que no tiene sentidos. Un cirujano podría cosernos un dedo dentro del cerebro y no lo sentiríamos. Toda la información que entra en nuestra mente lo hace como patrones espaciales y temporales en los axones.

¿Qué entiendo exactamente por patrones espaciales y temporales? Observemos uno por uno nuestros principales sentidos. La visión transporta información espacial y temporal. Los *patrones espaciales* son patrones coincidentes en el tiempo; se crean cuando múltiples receptores del mismo órgano sensorial se estimulan de forma simultánea. En la visión, el órgano sensorial es la retina. Entra una imagen en la pupila, es invertida por las lentes, alcanza la retina y crea un patrón espacial. Este patrón es enviado al cerebro. La gente tiende a pensar que es una pequeña foto invertida del mundo la que va a las áreas visuales, pero no es así como ocurre. No hay ninguna foto. Ha dejado de ser una imagen. En esencia, no es más que actividad eléctrica emitiendo patrones. Sus cualidades de imagen se pierden deprisa cuando la corteza cerebral maneja la información, pasando los componentes del patrón arriba y abajo entre las diferentes áreas, cambiándolos y filtrándolos.

La visión también se basa en *patrones temporales*, lo que significa que los patrones que entran en los ojos cambian constantemente a lo largo del tiempo. Pero mientras que el aspecto espacial de la visión resulta obvio por intuición, su aspecto temporal es menos evidente. Unas tres veces por segundo los ojos hacen un movimiento repentino. Se fijan en un punto y luego de improviso saltan a otro. Cada vez que los ojos se mueven, la imagen de la retina cambia, lo cual significa que los patrones transportados al cerebro también cambian por completo con cada sacudida ocular. Y así ocurre en el caso más simple posible, cuando miramos sentados una escena que no cambia. En la vida real, movemos constantemente la cabeza y el cuerpo, y caminamos por entornos que varían de forma continua. Nuestra impresión consciente es que hay un mundo estable lleno de objetos y gente a los que resulta fácil seguir. Pero esta impresión solo la hace posible nuestra capacidad cerebral para manejar un torrente de imágenes retinianas que nunca repiten un patrón exacto. La visión natural, experimentada como patrones que entran en el cerebro, fluye como un río. La visión se parece más a una canción que a una pintura.

Muchos investigadores de la visión ignoran las sacudidas oculares y los patrones en cambio constante de la vista. Trabajando con animales anestesiados, estudian cómo actúa la visión cuando un animal inconsciente la fija en un punto. Al hacerlo, están despreciando la dimensión temporal. En principio no tiene nada de malo; eliminar variables es un elemento central del método científico. Pero están desechando un componente crucial de la visión, lo que realmente la constituye. El

tiempo debe ocupar un lugar central en una explicación neurocientífica de la visión.

En cuanto al oído, estamos acostumbrados a pensar en el aspecto temporal del sonido. Es evidente por intuición que los sonidos, el lenguaje hablado y la música cambian con el tiempo. No se puede escuchar una canción completa al mismo tiempo, del mismo modo que no es posible oír una oración hablada en un instante. Una canción solo existe a lo largo del tiempo. Pero no siempre pensamos en los sonidos como un modelo espacial. En cierto modo, es lo contrario del caso de la visión: el aspecto temporal resulta evidente de inmediato, mas su aspecto espacial es menos obvio.

El oído también tiene un componente espacial. Convertimos los sonidos en potenciales de acciones mediante un órgano enrollado en cada oreja llamado caracol. Diminuto, opaco, con forma de espiral e insertado en el hueso más duro del cuerpo, el hueso temporal, el caracol, fue descifrado hace más de medio siglo por un físico húngaro, Georg von Beksey. Construyendo modelos del oído interno, Von Beksey descubrió que cada componente del sonido que oímos hace que vibre una parte diferente del caracol. Los tonos de frecuencia alta provocan vibraciones en la base rígida del caracol. Los tonos de frecuencia baja causan vibraciones en su parte más flexible y ancha. Los tonos de frecuencia media hacen vibrar los segmentos intermedios. Cada lugar del caracol está salpicado de neuronas que se estimulan cuando se las agita. En la vida cotidiana, a nuestro caracol le hacen vibrar grandes cantidades de frecuencias simultáneas todo el tiempo. Así pues, en cada momento hay un nuevo patrón espacial de estimulación por toda la extensión del caracol; en cada momento un nuevo patrón espacial fluye hasta el nervio auditivo. De nuevo, vemos que esta información sensorial se convierte en patrones espacio-temporales.

La gente no suele pensar que el tacto es un fenómeno temporal, pero está tan basado en el tiempo como en el espacio. Se puede efectuar un experimento para comprobarlo. Pídale a un amigo que ahueque la mano con la palma para arriba y cierre los ojos. Coloque un pequeño objeto ordinario en la palma —un anillo, un borrador, cualquier cosa valdrá— y solicítele que lo identifique sin mover ninguna parte de la mano. No tendrá más pista que el peso y tal vez el tamaño bruto. Luego dígame que mantenga los ojos cerrados y mueva los dedos sobre el objeto. Es muy probable que lo identifique de inmediato. Al permitir que los dedos se muevan, ha añadido tiempo a la percepción sensorial del tacto. Existe una analogía directa entre la fovea del centro de la retina y las puntas de los dedos, pues ambas poseen una gran precisión. Así pues, también el tacto es como una canción. Nuestra capacidad para hacer un uso complejo del tacto, como abotonarnos la camisa o abrir el cerrojo de la puerta delantera en la oscuridad, depende de patrones del sentido del tacto que varían constantemente con el tiempo.

Enseñamos a nuestros hijos que los humanos gozan de cinco sentidos: vista, oído, tacto, olfato y gusto, pero en realidad tenemos más. La visión es más bien tres sentidos: movimiento, color y luminancia (contraste de blanco y negro). El tacto tiene presión, temperatura, dolor y vibración. También contamos con un sistema completo de sensores que nos informan sobre nuestros ángulos de unión y posición corporal. Se denomina sistema propioceptivo (*propio* tiene la misma raíz latina que *propietario* y *propiedad*). No nos podríamos mover sin él. También disponemos del sistema vestibular del oído interno, que nos proporciona el sentido del equilibrio. Algunos de estos sentidos son más ricos y evidentes que otros, pero todos entran en nuestro cerebro como una corriente de patrones espaciales que fluyen a través del tiempo en los axones.

En realidad, nuestra corteza cerebral no conoce ni siente el mundo de forma directa. Lo único que conoce es el patrón que llega en los axones de entrada. Nuestra visión percibida del mundo se crea con estos patrones, incluido nuestro sentimiento de uno mismo. De hecho, nuestro cerebro no puede conocer de forma directa dónde termina nuestro cuerpo y comienza el mundo. Los neurocientíficos que estudian la imagen corporal han descubierto que nuestro sentido del *yo* es mucho más flexible de lo que parece. Por ejemplo, si yo le doy un rastrillo pequeño y le digo que lo use para alcanzar y coger cosas en lugar de emplear la mano, pronto sentirá que se ha convertido en parte de su cuerpo. Su cerebro cambiará sus expectativas para acomodarse a los nuevos patrones de entrada táctil. El rastrillo es incorporado literalmente a su mapa corporal.

\* \* \*

La idea de que los patrones de diferentes sentidos son equivalentes dentro de nuestro cerebro es bastante sorprendente y, aunque se entiende bien, sigue sin apreciarse por completo. Pongamos más ejemplos. El primero se puede reproducir en casa. Todo lo que se necesita es un amigo, una pantalla de cartulina y una mano de mentira. Para realizar por primera vez este experimento sería ideal contar con una mano de goma de las que se pueden comprar en una tienda de artículos de broma, pero también serviría una mano pintada en una hoja blanca de papel. Coloque su mano real sobre una mesa a unos cuantos centímetros de la falsa y sitúelas del mismo modo (con las puntas de los dedos señalando en la misma dirección y ambas palmas hacia arriba o hacia abajo). Luego ponga la pantalla entre las dos manos de modo que solo le permita ver la mano falsa. Mientras mira fijamente la mano falsa, la labor de su amigo consistirá en golpear ambas manos en puntos correspondientes. Por ejemplo, su amigo puede golpear ambos meñiques desde el nudillo hasta la uña a la misma velocidad, luego dar tres rápidos toques en la segunda articulación de ambos dedos índices con el mismo ritmo, a continuación describir unos cuantos círculos ligeros en el dorso de cada mano, y así sucesivamente. Transcurrido poco tiempo, las áreas de su cerebro a las que llegan juntos los patrones visuales y somatosensoriales —una de las áreas

de asociación que ya he mencionado en este capítulo— acaban confundidas. Sentirá las sensaciones correspondientes a la mano de goma como si fueran suyas.

Otro ejemplo fascinante de esta “equivalencia de patrones” se denomina sustitución sensorial. Puede revolucionar la vida de la gente que perdió la vista en la infancia, y tal vez algún día sea de gran ayuda para las personas que nacen ciegas. También podría producir nuevas tecnologías de interfaz para fabricar máquinas útiles para el resto de nosotros.

Paul Bach y Rita, profesor de ingeniería biomédica en la Universidad de Wisconsin, se ha dado cuenta de que en el cerebro todo consiste en patrones y ha desarrollado un método para detectar patrones visuales en la lengua humana. Usando este aparato de visualización, las personas ciegas están aprendiendo a “ver” a través de las sensaciones de la lengua.

Así es como funciona. El sujeto lleva una pequeña cámara en la frente. Las imágenes visuales se trasladan píxel a píxel a puntos de presión en la lengua. Una escena visual que se puede traducir como cientos de píxeles sobre una pantalla de televisión ordinaria se convierte en un patrón de diminutos puntos de presión sobre la lengua. El cerebro aprende enseguida a interpretar bien los patrones.

Una de las primeras personas en llevar el aparato montado en la lengua es Erik Weihenmayer, atleta de categoría mundial que se quedó ciego a los trece años y que da muchas conferencias sosteniendo que no va a permitir que la ceguera detenga sus ambiciones. En 2002, Weihenmayer coronó el Everest y se convirtió en la primera persona ciega que había intentado, y ya no digamos alcanzado, esa meta.

En 2003 probó la unidad colocada en la lengua y vio imágenes por primera vez desde su niñez. Fue capaz de distinguir una pelota que rodaba en el suelo hacia él, coger un refresco de la mesa y jugar a “piedra, papel, tijeras”. Después caminó por un pasillo, vio las aberturas de las puertas, examinó una puerta y su marco, y señaló que había una indicación en él. Las imágenes experimentadas al inicio como sensaciones en la lengua pronto pasaron a percibirse como imágenes en el espacio.

Estos ejemplos muestran una vez más que la corteza cerebral es extremadamente flexible y que las entradas al cerebro solo son patrones. No importa de dónde provienen estos; siempre que tengan una correlación temporal coherente, el cerebro puede hallarles sentido.

\* \* \*

**T**odo esto no debe resultar demasiado sorprendente si adoptamos el planteamiento de que el cerebro solo conoce patrones. Los cerebros son máquinas

de patrones. No es erróneo expresar las funciones cerebrales en términos de oído o visión, pero en el nivel más fundamental los patrones son la esencia del juego. Por muy diferentes que puedan parecer las actividades de varias áreas corticales, funciona el mismo algoritmo cortical básico. A la corteza cerebral no le importa si los patrones se originaron en la visión, el oído u otro sentido. No le importa si sus entradas provienen de un único órgano sensorial o de cuatro. Ni tampoco le importaría si diera la casualidad de que percibiéramos el mundo con sonar, radar o campos magnéticos, o si tuviéramos tentáculos en lugar de manos, o incluso si viviéramos en un mundo de cuatro dimensiones en lugar de tres.

Ello significa que no necesitamos ninguno de los sentidos o ninguna combinación particular de sentidos para ser inteligentes. Helen Keller no tenía vista ni oído, pero aprendió el lenguaje y se convirtió en una escritora más capacitada que la mayoría de la gente que ve y oye. Era una persona muy inteligente sin dos de nuestros principales sentidos, pero la increíble flexibilidad del cerebro le permitió percibir y comprender el mundo como lo hacen las personas con los cinco sentidos.

Este tipo de flexibilidad notable en la mente humana me proporciona grandes esperanzas acerca de la tecnología basada en el cerebro que crearemos. Cuando pienso en construir máquinas inteligentes, me pregunto por qué debemos limitarnos a nuestros sentidos conocidos. En cuanto podamos descifrar el algoritmo de la corteza cerebral y elaborar una ciencia sobre los patrones, podemos aplicarla a cualquier sistema que queramos hacer inteligente. Y uno de los grandes rasgos del sistema de circuitos inspirado en la corteza cerebral es que no necesitaremos ser especialmente inteligentes para programarlo. Del mismo modo que la corteza cerebral auditiva puede convertirse en "visual" en un hurón reconectado, y del mismo modo que la corteza cerebral visual encuentra un uso alternativo en las personas ciegas, un sistema que lleve el algoritmo cortical será inteligente basándose en cualquier tipo de patrones que decidamos darle. Sin embargo, sí es preciso que seamos inteligentes para organizar los amplios parámetros del sistema, y será necesario entrenarlo y educarlo. Pero los miles de millones de detalles neuronales que toman parte en la capacidad del cerebro de tener pensamientos complejos y creativos se ocuparán de sí mismos de forma tan natural como lo hacen en nuestros niños.

Por último, la idea de que los patrones son la moneda fundamental de la inteligencia conduce a algunas cuestiones filosóficas interesantes. Cuando me siento en una habitación con mis amigos, ¿cómo sé que están ahí, o incluso si son reales? Mi cerebro percibe un conjunto de patrones que concuerdan con otros que he experimentado en el pasado. Estos patrones corresponden a personas que conozco, sus rostros, sus voces, su conducta habitual, y todo tipo de datos sobre ellos. He aprendido a esperar que estos patrones ocurran juntos de formas predecibles. Pero cuando llegas a ello, no se trata más que de un modelo. Todo nuestro conocimiento del mundo es un modelo basado en patrones. ¿Estamos



seguros de que el mundo es real? Resulta divertido y extraño pensar en ello. Varios libros y películas de ciencia ficción exploran este tema. No se trata de afirmar que la gente o los objetos no estén ahí. Están realmente ahí. Pero nuestra certeza de la existencia del mundo se basa en la coherencia de los patrones y en cómo los interpretemos. La percepción directa no existe. No tenemos un sensor de "personas". Recordemos que el cerebro está en una caja oscura y silenciosa, sin ningún conocimiento más que los patrones que fluyen a lo largo del tiempo en sus fibras de entrada. Nuestra percepción del mundo se crea a partir de esos patrones, nada más. La existencia puede ser objetiva, pero los patrones espaciotemporales que fluyen en los manojos de axones de nuestro cerebro es todo lo que tenemos para seguir adelante.

Esta exposición pone de relieve la relación a veces cuestionada entre alucinación y realidad. Si se pueden alucinar sensaciones provenientes de una mano de goma y se puede "ver" a través de la estimulación del tacto de la lengua, ¿estamos siendo igualmente "engañados" cuando sentimos el tacto en nuestra propia mano o vemos con nuestros ojos? ¿Podemos confiar en que el mundo es como parece? Sí. El mundo existe de una forma absoluta muy próxima a como lo percibimos. Sin embargo, nuestros cerebros no pueden conocer el mundo absoluto de modo directo.

El cerebro sabe del mundo a través de un conjunto de sentidos que solo pueden detectar partes del mundo absoluto. Los sentidos crean patrones que son enviados a la corteza cerebral y procesados por el mismo algoritmo cortical para crear un modelo del mundo. De este modo, el lenguaje hablado y el lenguaje escrito se perciben de forma muy similar, pese a ser completamente diferentes a nivel sensorial. Asimismo, el modelo de Helen Keller del mundo estaba muy próximo al suyo y al mío, a pesar del hecho de que poseía un conjunto de sentidos muy reducido. Mediante estos patrones la corteza cerebral construye un modelo del mundo que se acerca a la cosa real, y luego, sorprendentemente, lo memoriza. La memoria, lo que sucede a estos patrones una vez que entran en la corteza cerebral, será el tema del siguiente capítulo.

## 4

### La Memoria

**Cuando** leen este libro, caminan por una calle llena de gente, escuchan una sinfonía o consuelan a un niño que llora, su cerebro es inundado con los patrones espaciales y temporales provenientes de todos sus sentidos. El mundo es un océano de patrones en cambio constante que llega a romper a nuestros cerebros. ¿Cómo logramos hallar sentido a esa avalancha? Los patrones llegan, pasan por varias partes del cerebro antiguo y acaban en la corteza cerebral. ¿Pero qué les sucede cuando entran en ella?

Desde los albores de la Revolución Industrial, la gente ha considerado al cerebro una especie de máquina. Sabía que no había engranajes ni dientes, pero era la mejor metáfora de que disponía. La información entraba en el cerebro de algún modo y la máquina-cerebro determinaba cómo debía reaccionar el cuerpo. Durante la era informática, el cerebro se ha considerado un tipo de máquina particular, el ordenador programable. Y, como hemos visto en el primer capítulo, los investigadores de la inteligencia artificial se han aferrado a esta postura, sosteniendo que su falta de avance se debe solo a lo pequeños y lentos que continúan siendo los ordenadores en comparación con el cerebro humano. Los ordenadores actuales solo equivalen al cerebro de un cocodrilo, afirman, pero cuando los fabriquemos más grandes y rápidos serán inteligentes como humanos.

En esta analogía del cerebro como ordenador hay un problema en buena medida ignorado. Las neuronas son bastante lentas comparadas con los transistores de un ordenador. Una neurona reúne entradas de sus sinapsis y las combina para decidir cuándo enviar un impulso a otra neurona. Una neurona normal puede hacer esto y resetearse en unas cinco milésimas de segundo, o alrededor de doscientas veces por segundo. Quizá parezca rápido, pero un ordenador moderno de silicio puede realizar mil millones de operaciones en un segundo, lo cual significa que una operación informática básica es cinco millones de veces más rápida que la operación más elemental de nuestro cerebro. Se trata de una diferencia grandísima. Así que ¿cómo es posible que un cerebro pueda ser más rápido y potente que nuestros ordenadores digitales más veloces? "No hay problema —dice la gente que opina que el cerebro es un ordenador—. El cerebro es un ordenador paralelo. Posee miles de millones de células computando todas al mismo tiempo. Este paralelismo multiplica con creces el poder de procesamiento del cerebro humano."

Siempre me ha parecido que este argumento es una falacia, y un sencillo experimento mental muestra por qué. Se denomina la "regla de cien pasos". Un humano puede realizar tareas considerables en mucho menos tiempo que un segundo. Por ejemplo, podría mostrarles una fotografía y pedirles que indicaran si hay un gato en la imagen. Su labor sería pulsar un botón si hay un gato, pero no

hacerlo si ven un oso, un jabalí o un nabo. Esta tarea es difícil o imposible de realizar para un ordenador actual, pero un humano puede hacerlo de forma fiable en medio segundo o menos. Pero las neuronas son lentas, así que en ese medio segundo la información que entra en nuestro cerebro solo es capaz de atravesar una cadena de cien neuronas. Es decir, el cerebro "computa" soluciones a problemas en cien pasos o menos, prescindiendo de cuántas neuronas puedan participar en total. Desde el instante en que la luz entra en nuestro ojo hasta el momento en que pulsamos el botón, podría participar una cadena no más larga de cien neuronas. Un ordenador digital que intente resolver el mismo problema necesitaría miles de millones de pasos. Cien instrucciones informáticas apenas bastan para mover un solo carácter en la pantalla del ordenador, y ya no digamos para hacer algo interesante.

Pero si tengo muchos miles de neuronas trabajando juntas, ¿no se parecen a un ordenador paralelo? No. Los cerebros operan en paralelo y los ordenadores paralelos operan en paralelo, pero es lo único que tienen en común. Los ordenadores paralelos combinan muchos ordenadores rápidos para trabajar en grandes problemas, como calcular el clima de mañana. Para predecir el tiempo se tienen que computar las condiciones físicas en muchos puntos del planeta. Cada ordenador puede trabajar en una localización diferente al mismo tiempo. Pero aunque es posible que haya cientos o incluso miles de ordenadores trabajando en paralelo, cada uno de ellos en particular sigue necesitando realizar miles de millones o billones de pasos para ejecutar su tarea mayor. El ordenador paralelo concebible no es capaz de hacer nada útil en cien pasos por muy grande o rápido que sea.

Veamos una analogía. Supongamos que le pido que transporte cien bloques de piedra al otro lado de un desierto. Puede llevar las piedras de una en una, con lo cual necesitará un millón de pasos para cruzar el desierto. Se da cuenta de que tardará mucho en conseguirlo solo, así que contrata a cien trabajadores para que lo hagan en paralelo. La tarea ahora avanza cien veces más deprisa, pero sigue requiriendo un mínimo de un millón de pasos cruzar el desierto. Contratar más trabajadores —incluso mil— no proporcionaría una ventaja adicional. Por muchos trabajadores de los que disponga, el problema no puede resolverse en menos tiempo de lo que se tarda en andar un millón de pasos. Lo mismo ocurre en el caso de los ordenadores paralelos. Después de un punto, añadir más procesadores no supone ninguna diferencia. Un ordenador, por muchos procesadores que tenga y por muy rápido que vaya, no puede "computar" la respuesta a problemas difíciles en cien pasos.

Así pues, ¿cómo logra el cerebro realizar tareas difíciles en cien pasos que el ordenador paralelo mayor imaginable no es capaz de realizar en un millón o mil millones de pasos? La respuesta es que el cerebro no "computa" las respuestas a los problemas, sino que las recupera de la memoria. En esencia, las respuestas

están almacenadas en la memoria un largo tiempo. No se necesitan más que unos cuantos pasos para recuperar algo de la memoria. Las lentas neuronas no solo poseen la rapidez necesaria para hacerlo, sino que ellas mismas constituyen la memoria. La corteza cerebral entera es un sistema de memoria. No es en absoluto un ordenador.

\* \* \*

**P**ermítanme mostrarles mediante un ejemplo la diferencia entre *computar* la solución a un problema y *usar la memoria* para resolverlo. Consideremos la tarea de coger una pelota. Alguien le tira una pelota, la ve desplazándose hacia usted, y en menos de un segundo la atrapa en el aire. No parece algo muy difícil, hasta que trata de programar el brazo de un robot para que lo haga. Como muchos estudiantes licenciados han descubierto, resulta casi imposible. Cuando los ingenieros o científicos informáticos abordan este problema, primero tratan de determinar dónde estará la pelota cuando llegue al brazo. Este cálculo requiere resolver un conjunto de ecuaciones del tipo de las que se aprenden en física en el instituto. A continuación tienen que armonizarse todas las uniones del brazo robótico para que muevan la mano a la posición adecuada. Ello supone resolver otro conjunto de ecuaciones matemáticas más difíciles que las primeras. Por último, debe repetirse esta operación completa múltiples veces, pues a medida que la pelota se acerca, el robot obtiene mejor información sobre su localización y trayectoria. Si el robot espera para empezar a moverse a conocer con exactitud dónde llegará la pelota, será demasiado tarde para cogerla. Debe comenzar a avanzar para atraparla cuando apenas tiene sentido de su localización, e ir la ajustando una y otra vez mientras esta se acerca. Un ordenador requiere millones de pasos para resolver las numerosas ecuaciones matemáticas que requiere coger la pelota. Y aunque podría programarse para solucionar dicho problema, la regla de cien pasos nos indica que un cerebro lo resuelve de un modo diferente. Emplea la memoria.

¿Cómo se coge la pelota empleando la memoria? Nuestro cerebro posee una memoria almacenada de las órdenes musculares requeridas para lograrlo (junto con muchas otras conductas aprendidas). Cuando se lanza una pelota, ocurren tres cosas. Primero, se recupera de forma automática la memoria apropiada ante la visión de la pelota. Segundo, la memoria recuerda una secuencia temporal de órdenes musculares. Y tercero, la memoria recuperada se ajusta a las particularidades del momento, tales como la trayectoria presente de la pelota y la posición de nuestro cuerpo. La memoria de cómo atrapar una pelota no estaba programada en nuestro cerebro; la aprendimos a lo largo de años de práctica repetitiva, y nuestras neuronas la guardan, no la calculan.

Tal vez piensen: "Espere un minuto. Cada acción de atrapar una pelota es ligeramente diferente. Acaba de decir que la memoria recuperada se tiene que ajustar de forma continua para adaptarse a las variaciones de la localización en

cada lanzamiento particular... ¿No requiere eso resolver las mismas ecuaciones que estamos tratando de evitar?”. Puede que así parezca, pero la Naturaleza solucionó el problema de la variación de un modo diferente y muy inteligente. Como veremos más adelante en este mismo capítulo, la corteza cerebral crea lo que se denominan *representaciones invariables*, que se ocupan de las variaciones del mundo de forma automática. Una analogía útil podría ser imaginarse lo que sucede cuando te sientas en una cama de agua: las almohadas y las otras personas de la cama se ven todas empujadas de forma espontánea a establecer una nueva configuración. La cama no calcula a qué altura debe elevarse cada objeto; las propiedades físicas del agua y la funda de plástico del colchón se encargan del ajuste de manera automática. Como veremos en el capítulo siguiente, a grandes rasgos, el diseño de la corteza cerebral de seis capas hace algo similar con la información que fluye por ella.

\* \* \*

Así pues, la corteza cerebral no se asemeja a un ordenador, sea paralelo o de otro tipo. En lugar de calcular respuestas a los problemas, utiliza memorias almacenadas para resolver problemas y generar conductas. Los ordenadores también tienen memoria en la forma de unidades de disco duro y chips de memoria; sin embargo, hay cuatro atributos de la memoria cortical que son fundamentalmente diferentes de la memoria informática:

- **La corteza cerebral almacena secuencias de patrones.**
- **La corteza cerebral recuerda los patrones por autoasociación.**
- **La corteza cerebral almacena los patrones en una forma invariable.**
- **La corteza cerebral almacena los patrones en una jerarquía.**

Analizaremos las tres primeras diferencias en este capítulo; en el capítulo 3 ya he presentado el concepto de jerarquía en la corteza cerebral, y en el capítulo 6 describiré su significado y funcionamiento.

La próxima vez que cuente un relato, deténgase a reflexionar sobre el hecho de que solo es capaz de narrar un aspecto a un tiempo. No puede contarme todo lo que sucedió a la vez, por muy deprisa que hable o yo escuche. Necesita terminar una parte para pasar a la siguiente. Ello no se debe solo a que el lenguaje es consecutivo; la narración escrita, oral y visual transmite un relato de forma consecutiva. Es así porque el relato se almacena en su cabeza de forma secuencial y solo puede recordarse en la misma secuencia. No puede recordar la historia entera de una vez. De hecho, es casi imposible pensar en algo complejo que no sea una serie de acontecimientos o pensamientos.

Tal vez también se haya dado cuenta de que al contar un relato algunas personas no son capaces de llegar al quid de la cuestión. Parecen irse por las ramas con detalles nimios y secundarios, lo cual resulta irritante. Quieres gritar: "¡Al grano!". Pero están refiriendo el relato como les sucedió en el tiempo y no saben contarlos de ningún otro modo.

Otro ejemplo: me gustaría que ahora se imaginara su casa. Cierre los ojos y visualícela. En su imaginación, vaya a la puerta principal. Imagine su aspecto. Ábrala. Pase dentro. Ahora mire a su izquierda. ¿Qué ve? Mire a la derecha. ¿Qué hay? Vaya al cuarto de baño. ¿Qué hay a la derecha? ¿Qué hay a la izquierda? ¿Qué hay en el cajón superior derecho? ¿Qué artículos guarda en su ducha? Sabe todas estas cosas, además de miles más, y puede recordarlas con gran detalle. Estos recuerdos se almacenan en su corteza cerebral. Cabría decir que estas cosas forman parte de la memoria de su hogar. Pero no puede pensar en todas ellas a la vez. Sin duda, son recuerdos relacionados, pero no hay modo de acordarse de todos los detalles de una vez. Tiene una memoria completa de su casa; pero para recordarla ha de pasar por ella en segmentos consecutivos, de forma muy semejante a como la experimenta.

Todas las memorias son así. Se tiene que pasar por la secuencia temporal de las cosas tal y como se hacen. Un patrón (acercarse a la puerta) recuerda el siguiente (traspasar la puerta), que a su vez evoca el siguiente (dirigirse al salón o subir las escaleras), y así sucesivamente. Cada uno es una secuencia que se ha seguido antes. Por supuesto, haciendo un esfuerzo consciente yo puedo cambiar el orden al describirles mi casa. Puedo saltar del sótano al segundo piso si decido centrarme en artículos sin seguir un orden secuencial. No obstante, una vez que comienzo a describir la habitación u objeto elegido, vuelvo a repetir una secuencia. No existen pensamientos verdaderamente aleatorios. El recuerdo de la memoria sigue casi siempre una ruta de asociación.

Conocen el alfabeto. Traten de decirlo al revés. No pueden porque no suelen experimentarlo de ese modo. Si desean saber qué siente un niño que está aprendiendo el alfabeto, traten de decirlo al revés. A eso es a lo que se enfrentan, y es durísimo. Nuestra memoria del alfabeto es una sucesión de patrones. No hay nada guardado o recordado en un instante o en una orden arbitraria. Lo mismo sucede con los días de la semana, los meses del año, el número de teléfono e innumerables cosas más.

Nuestro recuerdo de las canciones constituye un gran ejemplo de secuencias temporales en la memoria. Piensen en una melodía que conozcan. Me gusta emplear *Somewhere over the Rainbow*, pero cualquiera servirá. No son capaces de imaginarse la canción completa de una vez; solo en una secuencia. Pueden comenzar por el principio o tal vez con el coro, y luego la van tocando, poniendo las notas una tras otra. No pueden recordar la canción al revés, del mismo modo que son incapaces de recordarla de golpe. La escucharon por primera vez mientras

se interpretaba a lo largo del tiempo y solo son capaces de recordarla del mismo modo que la aprendieron.

Esto también es aplicable a los recuerdos sensoriales de nivel sensorial muy bajo. Reflexionemos sobre nuestra memoria táctil para las texturas. Nuestra corteza cerebral posee recuerdos de la sensación que produce coger un puñado de grava, deslizar los dedos sobre terciopelo y pulsar una tecla de un piano. Estas memorias se basan en secuencias idénticas a las del alfabeto y las canciones; la única diferencia es que estas son más cortas, duran meras fracciones de segundo en lugar de varios segundos o minutos. Si yo entierro su mano en un cubo de grava mientras duerme, cuando se despierte no sabrá qué estaba tocando hasta que mueva los dedos. Su memoria de la textura táctil de la grava se basa en secuencias de patrones recogidos por las neuronas que perciben la presión y vibración en su piel. Estas secuencias son diferentes de las que recibiría si su mano estuviera enterrada en arena, bolas de gomaespuma u hojas secas. En cuanto flexione la mano, las raspaduras y el deslizamiento de los guijarros crearían las secuencias de patrones reveladoras de la grava y desencadenarían la memoria apropiada en su corteza cerebral somatosensorial.

La próxima vez que salga de la ducha preste atención a su modo de secarse con la toalla. Yo descubrí que lo hacía con casi la misma sucesión exacta de frotés, palmadas y posiciones corporales todas las veces. Y mediante un experimento placentero descubrí que mi esposa respeta asimismo un patrón semirrígido cuando sale de la ducha. Es probable que usted también lo haga. Si sigue una secuencia, trate de cambiarla. Puede hacerlo, pero necesita centrarse en ello. Si su atención se desvía, volverá a caer en su patrón acostumbrado.

Todas las memorias se almacenan en las conexiones sinápticas que hay entre las neuronas. Debido al gran número de cosas que hemos guardado en nuestra corteza cerebral y a que en un momento determinado no podemos recordar más que una fracción diminuta de los recuerdos almacenados, es lógico pensar que solo un número limitado de sinapsis y neuronas de nuestro cerebro desempeñen un papel activo a un tiempo en la recuperación de la memoria. Cuando comenzamos a recordar qué hay en nuestra casa, se activa un conjunto de neuronas, que después hace que se active otro conjunto de neuronas, y así sucesivamente. La corteza cerebral de un humano adulto posee una capacidad de memoria increíblemente grande. Pero aunque tenemos almacenadas tantas cosas, no podemos recordar más que unas cuantas al mismo tiempo, y solo siguiendo una secuencia de asociaciones.

Realicemos un divertido ejercicio. Trate de recordar detalles de su pasado, pormenores de dónde vivía, lugares que visitó y personas que conoció. He descubierto que siempre puedo recuperar recuerdos de cosas en las que no había pensado durante muchos años. Hay miles de recuerdos detallados almacenados en las sinapsis de nuestros cerebros que rara vez se usan. En un momento

determinado del tiempo solo recordamos una fracción diminuta de lo que sabemos. La mayor parte de la información espera ociosa los indicios apropiados que la evoquen.

Por regla general, la memoria de los ordenadores no almacena secuencias de patrones. Puede lograrse que lo haga empleando varios trucos de *software* (como cuando se guarda una canción en el ordenador), pero no es algo automático. En contraste, la corteza cerebral sí almacena secuencias de forma automática. Hacerlo constituye un aspecto inherente de su sistema de memoria.

\* \* \*

**P**asemos ahora a considerar el segundo rasgo clave de nuestra memoria: su naturaleza autoasociativa. Como hemos visto en el capítulo 2, el término significa solo que los patrones están asociados consigo mismos. Un sistema de memoria autoasociativa es aquel que puede recordar patrones completos cuando se le dan solo entradas parciales o distorsionadas. Puede funcionar tanto con patrones espaciales como temporales. Si vemos los pies de nuestro hijo sobresaliendo detrás de las cortinas, automáticamente adivinamos su forma íntegra. Completamos el patrón espacial con una versión parcial de él. O imaginemos que vemos a una persona esperando el autobús, pero solo logramos distinguirla en parte porque está algo tapada por un arbusto. Nuestro cerebro no se confunde. Nuestros ojos solo ven partes de un cuerpo, pero nuestro cerebro llena el resto, creando una percepción de una persona completa tan potente que tal vez ni siquiera nos demos cuenta de que es una inferencia.

También completamos patrones temporales. Si recuerda un pequeño detalle de algo que pasó hace mucho tiempo, la secuencia del recuerdo completo puede acudir a su mente. La famosa serie de novelas de Marcel Proust, *En busca del tiempo perdido*, comenzaba con el recuerdo de cómo olía una magdalena y avanzaba para extenderse por mil y pico páginas. Durante la conversación es frecuente que no escuchemos todas las palabras si nos encontramos en un entorno ruidoso. No hay problema. Nuestro cerebro suple lo que se nos ha escapado con lo que esperamos oír. Es un hecho sabido que no escuchamos todas las palabras que percibimos. Algunas personas completan las oraciones de otros en voz alta, pero en nuestras mentes todos lo hacemos constantemente. Y no solo el final de las oraciones, sino también la parte media y los comienzos. Las más de las veces no nos damos cuenta de que estamos completando patrones de forma continuada, pero es un rasgo ubicuo y fundamental del modo como se almacenan los recuerdos en la corteza cerebral. En cualquier momento una parte puede activar el todo. Esta es la esencia de las memorias autoasociativas.

Nuestra corteza cerebral es una compleja memoria autoasociativa biológica. Durante cada momento de estado consciente, cada región funcional está esperando vigilante la llegada de patrones o fragmentos de patrones conocidos.



Puede hallarse sumido en una profunda reflexión sobre algo, pero en el momento en que aparece su amiga, sus pensamientos cambian hacia ella. Este cambio no es algo que se decida. La mera aparición de su amiga obliga a su cerebro a comenzar a recordar patrones asociados con ella. Es inevitable. Tras una interrupción, es frecuente que tengamos que preguntar: "¿En qué estaba pensando?". La conversación durante una comida con amigos sigue una ruta tortuosa de asociaciones. La charla puede comenzar con los alimentos que tenemos delante, pero la ensalada evoca el recuerdo asociado de la que hizo nuestra madre en nuestra boda, lo que lleva al recuerdo de la boda de otro, que conduce al recuerdo de dónde fueron de viaje de novios, a los problemas políticos de esa parte del mundo, y así sucesivamente. Los pensamientos y recuerdos están ligados por asociación y, una vez más, rara vez surgen pensamientos aleatorios. Las entradas del cerebro se unen entre sí autoasociativamente, completando el presente, y se unen en autoasociación a lo que suele seguir a continuación. Denominamos *pensamiento* a esta cadena de recuerdos, y aunque su camino no es determinista, tampoco poseemos un control pleno al respecto.

\* \* \*

**P**asemos ahora al tercer atributo principal de la memoria de la corteza cerebral: cómo forma las denominadas representaciones invariables. En este capítulo me ocuparé de las ideas básicas de dichas representaciones, y en el capítulo 6, de los detalles sobre cómo la corteza cerebral las crea.

La memoria de un ordenador está diseñada para almacenar información de la forma exacta como se presenta. Si copia un programa de un CD a un disco duro, se copia cada bit con una fidelidad total. Un único error o discrepancia entre las dos copias podría ocasionar el fallo del programa. La memoria de la corteza cerebral es diferente. Nuestro cerebro no recuerda con exactitud lo que ve, escucha o siente. No recordamos las cosas con una fidelidad completa no porque la corteza cerebral y sus neuronas sean descuidadas o proclives al error, sino porque el cerebro recuerda las relaciones importantes en el mundo, independientes de los detalles. Observemos varios ejemplos para ilustrar este punto.

Como hemos visto en el capítulo 2, durante décadas ha habido sencillos modelos de memoria autoasociativa y, como acabo de describir, el cerebro se acuerda de los recuerdos de forma autoasociativa. Pero existe una gran diferencia entre las memorias autoasociativas construidas por los investigadores de las redes neuronales y las de la corteza cerebral. Las memorias autoasociativas artificiales no emplean representaciones invariables y, por lo tanto, fracasan en algunos aspectos muy básicos. Imaginemos que tengo una foto de un rostro formado por una gran acumulación de puntos blancos y negros. Esta foto es un patrón, y si poseo una memoria autoasociativa artificial, puedo almacenar muchas fotos de rostros en ella. Nuestra memoria autoasociativa artificial es sólida en el sentido de

que si le facilito medio rostro o solo un par de ojos reconocerá esa parte de la imagen y completará las partes que faltan correctamente. Este mismo experimento se ha realizado muchas veces. Sin embargo, si muevo cada punto de la foto cinco píxeles a la izquierda, la memoria falla por completo al reconocer el rostro. Para la memoria autoasociativa artificial se trata de un patrón totalmente nuevo porque ninguno de los píxeles entre el patrón guardado con anterioridad y el nuevo están alineados. Por supuesto, ni a ustedes ni a mí nos costaría ver que el patrón cambiado es el mismo rostro. Es probable que ni siquiera notáramos el cambio. Las memorias auto-asociativas artificiales no logran reconocer los patrones si se mueven, rotan, cambia la escala o transforman de alguno de otros mil modos, mientras que nuestro cerebro maneja esas variaciones con facilidad. ¿Cómo podemos percibir que algo es lo mismo o constante cuando los patrones de entrada que lo representan son nuevos y cambiantes? Analicemos otro ejemplo.

Es probable que en estos momentos tenga un libro en las manos. Cuando lo mueve, cambia la iluminación, se recoloca en la silla o fija los ojos en partes diferentes de la página, el patrón de luz que penetra en la retina varía por completo. La entrada visual que recibe es diferente un momento tras otro y jamás se repite. De hecho, podría sujetar este libro durante cien años y nunca sería exactamente el mismo el patrón de la retina y, por lo tanto, el patrón que entra en el cerebro. Sin embargo, ni por un instante tiene duda alguna de que está sosteniendo un libro, el mismo libro en realidad. Los patrones internos de su cerebro que representan "este libro" no cambian aunque los estímulos que le informan de ello estén en flujo constante. De ahí que empleemos el término *representación invariable* para hacer referencia a la representación interna del cerebro.

Para poner otro ejemplo, piense en la cara de una amiga. La reconoce cada vez que la ve. Sucede de forma automática en menos de un segundo. No importa si se encuentra a medio metro, un metro o al otro lado de la habitación. Cuando está cerca, su imagen ocupa la mayor parte de su retina. Cuando está lejos, su imagen ocupa una pequeña porción de esta. Puede estar frente a usted, vuelta un poco de lado o de perfil. Tal vez esté sonriendo, entrecerrando los ojos o bostezando. Puede que la vea con luz brillante, en sombra o bajo las luces fantasmagóricas de ángulos extraños. Su semblante puede aparecer en posiciones y variaciones incontables. Para cada una el patrón de luz que llega a su retina es único, pero en todos los casos sabe al instante que es a ella a quien está mirando.

Saltemos la tapa y miremos lo que pasa dentro de su cerebro para que realice esa hazaña asombrosa. Sabemos por experimentos que si monitorizamos la actividad de las neuronas del área de entrada visual de su corteza cerebral, llamada V1, el patrón de actividad es diferente para cada visión distinta de la cara de su amiga. Cada vez que la cara se mueve o sus ojos realizan una nueva fijación, el patrón de actividad en V1 cambia, del mismo modo que lo hace el de la retina. Sin embargo,

si monitorizamos la actividad de las células de su área de reconocimiento de caras —una región funcional que se encuentra varios pasos por encima de V1 en la jerarquía cortical—, descubrimos estabilidad. Es decir, algunos de los conjuntos de células del área de reconocimiento visual permanecen activos mientras el rostro de su amiga se encuentre dentro de su campo de visión (o incluso mientras sea evocado por sus ojos mentales), prescindiendo de su tamaño, posición, orientación, escala y expresión. Esta estabilidad en la actividad celular es una representación invariable.

Si reflexionamos al respecto, esta tarea parece demasiado sencilla para merecer ser considerada un problema. Es tan automática como respirar. Parece nimia porque no somos conscientes de que está sucediendo. Y, en cierto sentido, es nimia porque nuestros cerebros pueden resolverla muy deprisa (recuérdese la regla de los cien pasos). Sin embargo, el problema de comprender cómo nuestra corteza cerebral forma representaciones invariables continúa siendo uno de los mayores misterios de la ciencia. ¿Es tan difícil?, se preguntan. Tanto que nadie, ni siquiera usando los ordenadores más potentes del mundo, ha sido capaz de resolverlo. Y no es porque no se haya intentado.

La especulación sobre este problema posee un historial antiguo. Se remonta hasta Platón, hace veintitrés siglos. El ateniense se preguntaba cómo la gente era capaz de pensar y conocer el mundo. Señaló que los modelos del mundo real de cosas e ideas siempre son imperfectos y diferentes. Por ejemplo, tenemos el concepto de un círculo perfecto, pero nunca hemos visto uno en realidad. Todos los dibujos de círculos son imperfectos. Incluso si lo dibujo con un compás de geómetra, el denominado círculo es representado por una línea oscura, mientras que la circunferencia de un círculo verdadero no tiene grosor. ¿Cómo, entonces, se llega a adquirir el concepto de círculo perfecto? O por tomar un ejemplo más casero, pensemos sobre nuestro concepto de perros. Cualquier perro que vemos es diferente de todos los demás, y cada vez que vemos al mismo perro particular obtenemos una visión distinta de él. Todos los perros son diferentes y nunca se puede ver a ningún perro particular del mismo modo exacto dos veces. Sin embargo, todas nuestras diversas experiencias con perros son canalizadas en un concepto mental de "perro" que es estable para todos ellos. Lo cual causaba perplejidad a Platón. ¿Cómo es posible que aprendamos y apliquemos conceptos en este mundo de formas infinitamente variadas y sensaciones en cambio constante?

La solución del filósofo fue su famosa Teoría de las Formas. Llegó a la conclusión de que nuestras mentes más elevadas debían estar amarradas a algún plano trascendente de suprarrealidad, donde existían ideas fijas y estables (Formas, con F mayúscula) en una perfección atemporal. Nuestras almas provenían de ese lugar místico antes del nacimiento, decidió, que es donde aprendieron primero sobre las Formas. Una vez que nacemos, conservamos un conocimiento latente de ellas. El

aprendizaje y la comprensión ocurren porque las formas del mundo real nos recuerdan las Formas con las que se corresponden. Somos capaces de saber de círculos y perros porque ambos desencadenan los recuerdos de nuestras almas de los Círculos y Perros.

Desde una perspectiva moderna, resulta bastante descabellado. Pero si nos despojamos de la metafísica ampulosa, podemos ver que en realidad hablaba de la invariancia. Su sistema explicativo estaba completamente errado, pero su intuición de que esta era una de las cuestiones más importantes que podemos plantearnos sobre nuestra naturaleza dio en el blanco.

\* \* \*

**P**ara que no tengan la impresión de que la invariancia se reduce a la visión, examinemos algunos ejemplos de otros sentidos. Reflexionemos sobre el tacto. Cuando meten la mano en la guantera de su coche para buscar las gafas de sol, sus dedos no tienen más que rozarse con ellas para que sepan que las han encontrado. No importa qué parte de la mano establece el contacto; puede ser el pulgar, cualquier parte de un dedo cualquiera o la palma. Y el contacto puede ser con cualquier parte de las gafas, ya sea un cristal, el puente, la patilla o parte de la montura. Un segundo moviendo cualquier parte de la mano sobre cualquier porción de las gafas basta para que el cerebro las identifique. En cada caso, el flujo de patrones espaciales y temporales procedente de sus receptores de tacto es completamente diferente —distintas áreas de piel, diferentes partes del objeto—, pero coge las gafas sin pensarlo.

O analicemos la tarea sensomotora de meter la llave en el contacto del coche. La posición del asiento, el cuerpo, el brazo y la mano son levemente diferentes cada vez. A ustedes les parece la misma acción sencilla y repetitiva día tras día, pero es debido a que tienen una representación invariable de ella en el cerebro. Si trataran de hacer un robot que pudiera entrar en el coche y meter la llave, verían de inmediato lo difícilísimo que resulta a menos que se aseguren de que el robot esté en la misma posición exacta y sostenga la llave de la misma manera precisa cada vez. E incluso si fueran capaces de lograrlo, el robot necesitaría ser reprogramado para coches diferentes. A los programas de los robots y los ordenadores, al igual que a las memorias auto-asociativas artificiales, les cuesta muchísimo manejar la variación.

Otro ejemplo interesante es la firma. En algún lugar de la corteza cerebral motora, en el lóbulo frontal, tenemos una representación invariable de nuestro autógrafo. Cada vez que firmamos con nuestro nombre, usamos la misma secuencia de rasgos, ángulos y ritmos. Es así lo hagamos minuciosamente con un bolígrafo de punta fina, de manera ampulosa como John Hancock, en el aire con el codo, o toscamente con un lápiz sujeto entre los dedos del pie. Parece algo diferente cada vez, por supuesto, sobre todo en las condiciones extrañas que acabo de

mencionar. Sin embargo, prescindiendo de la escala, el instrumento de escritura o la combinación de partes corporales, siempre empleamos el mismo “programa motor” abstracto para hacerlo.

Por el ejemplo de la firma se puede ver que la representación invariable de la corteza cerebral motora es, en ciertos sentidos, la imagen reflejada de la representación invariable de la corteza cerebral sensorial. En la parte sensorial, una amplia variedad de patrones de entrada pueden activar un conjunto estable de células que representa algún patrón abstracto (la cara de nuestra amiga o nuestras gafas de sol). En la parte motora, un conjunto de células estable que representa alguna orden motora abstracta (atrapar una pelota, firmar nuestro nombre) es capaz de expresarse utilizando una amplia variedad de grupos musculares y respetando una extensa gama de otras limitaciones. Esta simetría entre percepción y acción es la que cabe esperar si, como propuso Mountcastle, la corteza cerebral ejecuta un único algoritmo básico en todas las áreas.

Volvamos a la corteza cerebral sensorial y reflexionemos de nuevo sobre la música para analizar un ejemplo final. (Me gusta emplear el recuerdo de la música como ejemplo porque es fácil observar todos los temas que la corteza cerebral debe resolver.) La representación invariable en la música se ilustra por nuestra capacidad para reconocer una melodía en cualquier clave. La clave en la que se toca una melodía hace referencia a la escala musical con la que se ha compuesto. La misma melodía tocada en diferentes claves comienza con notas distintas. Una vez que elegimos la clave para una interpretación, hemos determinado el resto de las notas de la melodía. Toda melodía puede ser tocada en cualquier clave, lo que significa que cada interpretación de la “misma” melodía en una nueva clave es en realidad una secuencia de notas totalmente diferente. Cada interpretación estimula un conjunto de localizaciones en el caracol completamente distinto, provocando que un conjunto de patrones espaciotemporales diferentes entren en la corteza cerebral auditiva... y, sin embargo, percibimos la misma melodía en cada caso. A menos que tengamos un oído perfecto, ni siquiera seremos capaces de distinguir la misma canción tocada en dos claves diferentes sin volver a escucharlas otra vez.

Pensemos en la canción *Somewhere over the Rainbow*. Es probable que la aprendieran al escuchar a Judy Garland cantarla en la película *El mago de Oz*, pero, a menos que gocen de un oído perfecto, no recordarán en qué clave lo hacía (la bemol). Si me siento al piano y comienzo a tocar la canción en una clave en la que jamás la hayan escuchado —digamos en re—, sonará igual. No se darán cuenta de que todas las notas son diferentes de las de la versión que conocen. Ello significa que su memoria de la canción debe estar en una forma que pasa por alto el tono. La memoria tiene que almacenar las relaciones importantes de la canción, no las notas reales. En este caso, las relaciones importantes son el tono relativo de las notas, o “intervalos”. *Somewhere over the Rainbow* comienza con una octava alta, seguida por un semitono bajo, una tercera mayor baja, y así sucesivamente.

La estructura de intervalos de la melodía es la misma para cualquier interpretación, sea cual fuere la clave. Su capacidad para reconocer la canción en cualquier clave indica que su cerebro la ha guardado en su forma invariable de tono.

De forma similar, la memoria del rostro de su amiga también debe estar almacenada en una forma que es independiente de cualquier visión particular. Lo que hace reconocible su cara son sus dimensiones relativas, sus colores relativos y sus proporciones relativas, no el aspecto que presentó durante un instante el martes pasado en la comida. Hay "intervalos espaciales" entre los rasgos de su rostro, del mismo modo que hay "intervalos de tono" entre las notas de una canción. Su cara es ancha en relación con sus ojos. Su nariz es pequeña en relación con la amplitud de sus ojos. El color de su pelo y el de sus ojos presentan una relación relativa similar que permanece constante aun cuando en condiciones de luz diferentes sus colores absolutos cambien mucho. Cuando memorizaron su rostro, memorizaron esos atributos relativos.

Creo que ocurre una abstracción de la forma, similar en toda la corteza cerebral, en cada una de sus regiones. Se trata de una propiedad general de aquella. Los recuerdos se almacenan en una forma que capta la esencia de las relaciones, no los detalles del momento. Cuando vemos, palpamos o escuchamos algo, la corteza cerebral toma la entrada detallada y muy específica para convertirla en una forma invariable, que es la que se guarda, y es con esta con la que se compara cada nuevo patrón de entrada. El almacenamiento de la memoria, su recuperación y reconocimiento tienen lugar en el plano de las formas invariantes. No existe un concepto equivalente en los ordenadores.

\* \* \*

**L**o cual saca a colación un interesante problema. En el capítulo siguiente sostengo que una función importante de la corteza cerebral es utilizar esta memoria para realizar predicciones. Pero puesto que la corteza cerebral almacena formas invariantes, ¿cómo puede efectuar predicciones específicas? Veamos algunos ejemplos para ilustrar el problema y la solución.

Imaginemos que es el año 1890 y estamos en un pueblo fronterizo del Oeste americano. Nuestra amada está tomando el tren desde el Este para reunirse con nosotros en nuestro nuevo hogar de la frontera. Por supuesto, queremos recibirla en la estación cuando llegue. Durante unas cuantas semanas antes del día de su venida, nos dedicamos a observar cuándo llegan los trenes y cuándo salen. No hay horario, y, hasta donde hemos podido concluir, el tren nunca entra ni se marcha a la misma hora durante el día. Comienza a parecer que no seremos capaces de predecir cuándo llegará su tren. Pero entonces nos damos cuenta de que existe cierta estructura en las llegadas y salidas de trenes. El tren procedente del Este llega cuatro horas después del que sale en esa dirección. Este espacio de cuatro horas es constante un día tras otro, aunque los tiempos específicos varíen mucho.

El día de su llegada, aguardamos la aparición del tren con destino al Este y, cuando lo vemos, ponemos en hora el reloj. Cuatro horas después vamos a la estación y nos encontramos con el tren de nuestra amada justo cuando llega. Esta parábola ilustra tanto el problema al que se enfrenta la corteza cerebral como la solución que emplea para resolverlo.

El mundo tal como lo ven nuestros sentidos nunca es el mismo; al igual que la hora de llegada y salida del tren, siempre es diferente. Comprendemos el mundo buscando una estructura invariable en el flujo de entradas en cambio constante. Sin embargo, esta estructura invariable no basta para emplearla como base para realizar predicciones específicas. Saber que el tren llega cuatro horas después de que ha salido no nos permite aparecer en el andén justo a tiempo para recibir a nuestra amada. Para realizar una predicción específica, el cerebro debe combinar el conocimiento de la estructura invariable con los detalles más recientes. Predecir la hora de llegada del tren requiere reconocer la estructura de cuatro horas de intervalo en su horario y combinarla con el conocimiento detallado de la hora en que salió el último tren con destino al Este.

Cuando escuchamos una canción conocida tocada al piano, nuestra corteza cerebral predice la siguiente nota antes de que suene. Pero la memoria de la canción, como hemos visto, está en una forma invariable de tono. Nuestra memoria nos indica cuál es el siguiente intervalo, pero no nos dice nada sobre la nota real. Para predecir la nota exacta que va a continuación es preciso combinar el intervalo siguiente con la última nota específica. Si el intervalo siguiente es una tercera mayor y la última nota que hemos escuchado fue do, cabe predecir que la siguiente nota específica será mi. Escuchamos en nuestra mente mi, no "tercera mayor". Y a menos que hayamos identificado mal la canción o el pianista se equivoque, nuestra predicción es acertada.

Al ver el rostro de nuestra amiga, nuestra corteza cerebral aporta y predice la miríada de detalles de su imagen única en ese instante. Comprueba que sus ojos son los cabales, y que su nariz, labios y pelo son como deben ser. Nuestra corteza cerebral efectúa estas predicciones con una gran especificidad. Es capaz de predecir detalles insignificantes sobre su rostro aun cuando jamás la hayamos visto antes en esa orientación o ambiente particular. Si sabemos a ciencia cierta dónde están los ojos y nariz de nuestra amiga, y conocemos la estructura de su rostro, podemos predecir con exactitud dónde deben estar sus labios. Si sabemos que su piel está teñida de naranja por la luz de la puesta de sol, sabemos de qué color debe aparecer su pelo. Una vez más, nuestro cerebro lo hace combinando la memoria de la estructura invariable de su rostro con los particulares de nuestra experiencia inmediata.

El ejemplo del horario del tren no es más que una analogía de lo que sucede en nuestra corteza cerebral, pero los ejemplos de la melodía y el rostro, no. La combinación de representaciones invariables y entradas inmediatas para establecer

predicciones detalladas es exactamente lo que sucede. Es un proceso ubicuo que ocurre en todas las regiones de la corteza cerebral. Así es como efectuamos predicciones específicas sobre la habitación en la que estamos sentados en este momento. Así es como somos capaces de predecir no solo las palabras que dirán los demás, sino también el tono de voz que emplearán para hacerlo, su acento y desde qué parte de la habitación esperamos escucharla. Así es como sabemos con precisión cuándo nuestro pie dará en el suelo y cómo será subir un tramo de escalones. Así es como somos capaces de firmar nuestro nombre con el pie, o atrapar una pelota que nos lanzan.

Las tres propiedades de la memoria cortical analizadas en este capítulo (almacenamiento de secuencias, memoria autoasociativa y representaciones invariables) son ingredientes necesarios para predecir el futuro basado en recuerdos del pasado. En el capítulo siguiente propongo que efectuar predicciones constituye la esencia de la inteligencia.



## Un Nuevo Marco para la Inteligencia

**Un** día de abril de 1986 me puse a reflexionar sobre qué significaba “entender” algo. Me había pasado meses luchando con la pregunta fundamental de qué hacen los cerebros cuando no generan conducta. ¿Qué hace un cerebro cuando escucha de forma pasiva una charla? ¿Qué hace ahora mismo su cerebro mientras está leyendo? La información entra en el cerebro, pero no sale. ¿Qué le sucede? Sus conductas en este momento son probablemente básicas —como la respiración y los movimientos oculares—, pero cuando se encuentra en estado consciente su cerebro hace mucho más que eso mientras usted lee y comprende estas palabras. Comprender debe ser el resultado de la actividad neuronal. ¿Pero cuál? ¿Qué hacen las neuronas cuando entienden?

Al mirar alrededor en mi despacho ese día vi objetos conocidos: sillas, carteles, ventanas, plantas, lápices y demás. Me rodeaban cientos de artículos. Mis ojos los veían mientras miraba en torno, pero el solo hecho de verlos no me incitaba a realizar una acción. No se invocaba o requería ninguna conducta, mas de algún modo yo “entendía” la habitación y su contenido. Estaba haciendo lo que no podía la habitación china de Searle, y sin tener que pasar nada a través de una ranura. Comprendía, pero no había ninguna acción que lo demostrara. ¿Qué significaba “comprender”?

Fue mientras cavilaba sobre este dilema cuando tuve una percepción, uno de esos intensos momentos emocionales en que de repente lo que era una maraña de confusión se vuelve claro y comprensible. Todo lo que hice fue plantearme qué pasaría si un nuevo objeto, algo que nunca hubiera visto antes, apareciera en la habitación, digamos una taza de café azul.

La respuesta parecía simple. Me daría cuenta de que el nuevo objeto no correspondía a ese lugar. Llamaría mi atención por ser nuevo. No necesitaría preguntarme de forma consciente si la taza de café era nueva. Se limitaría a resaltar como algo que pertenecía al lugar. En esta respuesta de apariencia trivial subyace un valioso concepto. Para apreciar que algo es diferente, ciertas neuronas de mi cerebro antes inactivas habrían entrado en funcionamiento. ¿Cómo sabrían dichas neuronas que la taza de café azul era nueva y los cientos de objetos restantes que había en la habitación no? La respuesta sigue sorprendiéndome. Nuestro cerebro emplea memorias almacenadas para realizar predicciones constantes sobre todo lo que vemos, sentimos y escuchamos. Cuando miro alrededor de la habitación, mi cerebro está utilizando recuerdos para formar predicciones sobre lo que espera experimentar antes de que suceda. La vasta mayoría de las predicciones ocurren sin que tengamos conciencia de ello. Es como si diferentes partes de mi cerebro estuvieran diciendo: “¿Está el ordenador en mitad del escritorio? Sí. ¿Es negro? Sí. ¿Está la lámpara en la esquina derecha del

escritorio? Sí. ¿Está el diccionario donde lo dejé? Sí. ¿Es la ventana rectangular y la pared vertical? Sí. ¿Entra la luz del sol desde la dirección adecuada para la hora del día? Sí.” Pero cuando entra un patrón visual que no había memorizado en ese contexto, la predicción se incumple y mi atención se dirige al error.

Por supuesto, el cerebro no habla consigo mismo mientras realiza predicciones, y no las hace en serie. Tampoco se limita a realizar predicciones sobre distintos objetos como tazas de café. Nuestro cerebro efectúa predicciones constantes sobre la misma estructura del mundo en que vivimos, y lo hace en paralelo. Estará igual de dispuesto para detectar una textura extraña, una nariz deformada o un movimiento inusual. El carácter omnipresente de estas predicciones, inconscientes en su mayoría, no se advierte a primera vista, motivo por el cual tal vez hemos pasado por alto su importancia durante tanto tiempo. Suceden de modo tan automático, con tanta facilidad, que no logramos desentrañar lo que está pasando dentro de nuestros cráneos. Espero lograr comunicarles la fuerza de esta idea. La predicción es tan dominante que lo que “percibimos” —es decir, cómo aparece ante nosotros el mundo— no proviene únicamente de nuestros sentidos. Lo que percibimos es una combinación de lo que apreciamos y de las predicciones de nuestro cerebro derivadas de la memoria.

\* \* \*

**M**inutos después concebí un experimento mental para ayudar a transmitir lo que había comprendido en ese momento. Lo denominé el experimento de la puerta modificada. Es como sigue.

Cuando usted llega a casa cada día, suele tardar unos cuantos segundos en traspasar la puerta principal o cualquier puerta que use. Llega hasta ella, gira el pomo, entra y la cierra detrás de usted. Es un hábito firmemente establecido, algo que hace de forma constante y al que presta escasa atención. Supongamos que mientras usted se encuentra fuera me introduzco en su casa y cambio algo de su puerta. Podría ser casi cualquier cosa. Podría desplazar el pomo un milímetro, cambiar un pomo redondo por otro alargado, o uno de latón por otro cromado. Podría variar el peso de la puerta, sustituyendo el roble macizo por contrachapado, o viceversa. Podría hacer que los goznes chirriaran o que se deslizaran sin fricción. Podría ampliar o reducir la anchura de la puerta y su marco. Podría cambiar su color, añadir una aldaba donde solía estar la mirilla, o una ventana. Soy capaz de imaginar miles de cambios que harían que su puerta fuese desconocida para usted. Cuando llegara a casa ese día e intentara abrirla, detectaría enseguida que algo va mal. Tal vez le llevaría unos segundos de reflexión darse cuenta exacta de qué es, pero percibiría el cambio con mucha rapidez. Cuando su mano alcance el pomo desplazado, notará que no está en la ubicación correcta. O cuando vea la nueva ventana de la puerta, algo le parecerá raro. O si ha cambiado su peso, no empujará con la fuerza adecuada y se sorprenderá. El asunto es que se dará cuenta de uno de miles de cambios en un lapso de tiempo muy corto.

¿Cómo lo hace? ¿Cómo percibe esos cambios? Los ingenieros informáticos o de la inteligencia artificial abordarían este problema creando una lista de todas las propiedades de la puerta y poniéndola en una base de datos, con campos para cada atributo que una puerta puede tener y entradas específicas para su puerta particular. Cuando usted se aproxime a la puerta, el ordenador indagaría en toda la base de datos, buscando anchura, color, tamaño, posición del pomo, peso, sonido y demás. Aunque en superficie esta operación pueda parecer similar al modo como he descrito que mi cerebro comprobaba cada una de su miríada de predicciones mientras miraba alrededor en mi despacho, la diferencia es real y de largo alcance. La estrategia de la inteligencia artificial es inverosímil. En primer lugar, es imposible especificar por adelantado todos los atributos que una puerta puede tener. La lista es en potencia interminable. En segundo lugar, necesitaríamos contar con listas similares para cada objeto que nos encontremos cada segundo de nuestras vidas. En tercer lugar, nada de lo que sabemos sobre cerebros y neuronas sugiere que sea así como funcionan. Y, por último, las neuronas son demasiado lentas para aplicar bases de datos del tipo que emplean los ordenadores. Tardaría veinte minutos en lugar de dos segundos en darse cuenta de los cambios cuando pasa por la puerta.

Solo hay un modo de interpretar su reacción ante la puerta modificada: su cerebro realiza predicciones sensoriales de bajo nivel sobre lo que espera ver, escuchar y sentir en cada momento determinado, y lo hace en paralelo. Todas las regiones de su corteza cerebral tratan de predecir a la vez cuál será su próxima experiencia. Las áreas visuales efectúan predicciones sobre bordes, formas, objetos, ubicaciones y movimientos; las áreas auditivas, sobre tonos, dirección de la fuente y patrones de sonido; y las áreas somatosensoriales, sobre tacto, textura, contorno y temperatura.

La "predicción" significa que las neuronas que participan en la apreciación de la puerta se activan antes de recibir la entrada sensorial; cuando llega esta, se compara con lo que se esperaba. Mientras se acerca a la puerta, su corteza cerebral va formando un montón de predicciones basadas en la experiencia pasada. Cuando llega a ella, predice qué sentirá en los dedos cuando la toque y en qué ángulo estarán sus articulaciones cuando lo haga. Cuando comienza a empujarla para abrirla, su corteza cerebral predice cuánta resistencia ofrecerá y cómo sonará. Cuando se cumplen todas sus predicciones, traspasará el umbral sin tener conciencia de que dichas predicciones se han verificado. Pero si sus expectativas se vulneran, el error hará que se dé cuenta. Las predicciones acertadas dan como resultado la comprensión. La puerta está normal. Las predicciones erróneas inducen a confusión y suscitan su atención. El pomo de la puerta no está donde debería. La puerta es demasiado ligera. La puerta está descentrada. La textura del pomo no es la adecuada. Realizamos predicciones continuas de bajo nivel en paralelo con todos nuestros sentidos.

Pero eso no es todo. Mi propuesta es mucho más ambiciosa. La predicción no es solo una de las cosas que hace nuestro cerebro. Es la función primordial de la corteza cerebral y la base de la inteligencia. La corteza cerebral es un órgano de predicción. Si queremos entender qué es la inteligencia, qué es la creatividad, cómo funciona nuestro cerebro y cómo construir máquinas inteligentes, debemos comprender la naturaleza de estas predicciones y cómo las realiza la corteza cerebral. Hasta la conducta se entiende mejor como un producto derivado de la predicción.

\* \* \*

**No** sé cuál fue la primera persona que sugirió que la predicción es clave para comprender la inteligencia. En la ciencia y en la industria, nadie inventa nada completamente nuevo; se trata más bien de amoldar ideas existentes a nuevos marcos. Los componentes de una nueva idea suelen estar flotando en el entorno del discurso científico antes de su descubrimiento; lo que suele ser nuevo es la cohesión de esos componentes en un todo abarcador. De igual modo, la idea de que la función primordial de la corteza cerebral es realizar predicciones no es completamente nueva. Ha estado flotando alrededor en varias formas durante algún tiempo, pero todavía no ha asumido su posición legítima en el centro de la teoría del cerebro y la definición de la inteligencia.

Resulta irónico que algunos de los pioneros de la inteligencia artificial tuvieran la noción de que los ordenadores debían construir un modelo del mundo y utilizarlo para realizar predicciones. En 1956, por ejemplo, D. M. Mackay sostuvo que las máquinas inteligentes deberían tener un “mecanismo de respuesta interno” diseñado para “cotejar lo que se recibe”. No empleó las palabras “memoria” y “predicción”, pero su pensamiento iba en esa línea.

Desde mediados de la década de 1990, términos como *inferencia*, *modelos generativos* y *predicción* han entrado en el vocabulario científico. Todos se refieren a ideas relacionadas. Como ejemplo, en su libro de 2001, *El cerebro y el mito del yo*, Rodolfo Llinás, de la Escuela de Medicina de la Universidad de Nueva York, escribió: “La capacidad para predecir el resultado de acontecimientos futuros — crucial para que el movimiento tenga éxito— es muy probable que sea la función primordial y más común de todas las funciones globales del cerebro”. Científicos como David Mumford de la Universidad Brown, Rajesh Rao de la Universidad de Washington, Stephen Grossberg de la Universidad de Boston y muchos más han escrito y teorizado sobre el papel de la realimentación y la predicción de varios modos. Hay un subcampo completo en la matemática dedicado a las redes bayesianas. Bautizadas de ese modo por Thomas Bayes, pastor inglés nacido en 1702 que fue pionero en estadística, las redes bayesianas utilizan la teoría de la probabilidad para efectuar predicciones.

Lo que ha faltado es colocar estos elementos dispares dentro de un marco teórico coherente. Sostengo que no se ha hecho antes, y constituye la meta de este libro.

\* \* \*

**A**ntes de entrar en detalles sobre cómo realiza predicciones la corteza cerebral, analicemos algunos ejemplos adicionales. Cuanto más piensen en esta idea, más se darán cuenta de que la predicción es omnipresente y la base de nuestro modo de comprender el mundo.

Esta mañana hice tortitas. En un punto del proceso, alargué la mano por debajo de la encimera para abrir la puerta de un armario. Por intuición sabía sin verlo cuál sería el tacto —en este caso, el del tirador del armario— y cuándo lo sentiría. Retorcí el extremo del envase de leche esperando que se doblaría y luego desprendería. Encendí la plancha esperando presionar el mando un poco y luego girarlo con cierta resistencia. Esperaba escuchar el suave sonido de la llama de gas más o menos un segundo después. Cada minuto pasado en la cocina realicé docenas o cientos de movimientos, y cada uno suponía muchas predicciones. Lo sé porque si alguno de estos movimientos comunes hubiera tenido un resultado diferente del esperado me habría dado cuenta.

Cada vez que pone el pie en el suelo mientras camina, su cerebro predice cuándo se detendrá y cuánta “elasticidad” tendrá el material que pisa. Si alguna vez le ha fallado un escalón en un tramo de escaleras, sabe con cuánta rapidez se da cuenta de que algo va mal. Baja el pie y en el momento en que “pasa a través” del escalón previsto sabe que tiene dificultades. El pie no aprecia nada, pero su cerebro hizo una predicción que no se cumplió. Un robot dirigido por ordenador se caería sin más al no darse cuenta de que pasaba algo, mientras que usted lo sabría tan pronto como su pie sobrepasara en una fracción de milímetro el lugar donde su cerebro ha esperado que se detenga.

Cuando escucha una melodía conocida, oye la nota siguiente en su cabeza antes de que suene. Cuando escucha un álbum que le gusta, oye el comienzo de cada canción un par de segundos antes de que se inicie. ¿Qué sucede? Las neuronas de su cerebro que se estimularán cuando escuche la nota siguiente lo hacen por adelantado y, de este modo, “escucha” la canción en su cabeza. Las neuronas se estimulan en respuesta a la memoria, que puede ser muy duradera. No es inusual escuchar un álbum de música por primera vez en muchos años y seguir oyendo la canción siguiente de forma automática antes de que la anterior haya acabado. Y se crea una agradable sensación de leve incertidumbre cuando se escucha el álbum favorito sonando al azar, pues se sabe que la predicción de la canción siguiente errará.

Cuando escuchamos hablar a alguien, a menudo sabemos lo que va a decir antes de que haya terminado, o al menos pensamos que lo sabemos. A veces ni siquiera

escuchamos lo que dice el orador en realidad, sino lo que esperamos escuchar. (Esto me sucedía con tanta frecuencia cuando era pequeño que mi madre me llevó al médico para que me revisara el oído.) Y es así porque la gente tiende a emplear expresiones comunes en buena parte de su conversación. Si yo digo: "*How now brown...*"; su cerebro activará las neuronas que representan la palabra *cow* antes de que la pronuncie (aunque si el inglés no es su lengua nativa, tal vez no tenga idea de lo que estoy hablando. Se trata de un conocido juego de palabras que significa qué pasa ahora). Por supuesto, no sabemos todas las veces lo que van a decir los demás. La predicción no siempre es exacta, pues nuestras mentes funcionan realizando predicciones probables sobre lo que está a punto de suceder. A veces conocemos con exactitud lo que va a ocurrir, otras veces nuestras expectativas se distribuyen entre varias posibilidades. Si estuviéramos hablando en una mesa durante una comida y yo dijera: "Por favor, pásame la...", su cerebro no se sorprendería si añadiera a continuación "sal", "pimienta" o "mostaza". En cierto sentido, su cerebro predice todos estos resultados posibles al mismo tiempo. Sin embargo, si dijera: "Por favor, pásame la acera", usted sabría que algo va mal.

Volviendo a la música, en ella también podemos ver predicciones probables. Si estamos escuchando una canción que jamás hemos oído antes, de todos modos podemos tener previsiones bastante sólidas. En la música *country* espero un compás regular, un ritmo repetido, frases que duran el mismo número de compases y canciones que terminen en tono de altura. Tal vez no sepan qué significan estos términos, pero —suponiendo que hayan escuchado música similar— su cerebro predice de forma automática compases, ritmos repetidos, terminación de las frases y finales de las canciones. Si una nueva canción viola estos principios, saben de inmediato que algo va mal. Piensen en ello durante un segundo. Escuchan una canción que jamás han oído, su cerebro experimenta un patrón que jamás ha experimentado, y sin embargo realizan predicciones y pueden decir si algo va mal. La base de estas predicciones en buena parte inconscientes es el conjunto de memorias que están almacenadas en nuestra corteza cerebral. Nuestro cerebro no puede afirmar con exactitud qué pasará a continuación, pero sí predice qué patrón de nota es probable que vaya y cuál no.

Todos hemos tenido la experiencia de darnos cuenta de improviso de que una fuente de ruido de fondo constante, como una perforadora de mano o una excavadora, acaba de cesar; sin embargo, no habíamos percibido el sonido mientras ocurría. Nuestras áreas auditivas estaban prediciendo su continuación un momento tras otro, y mientras el sonido no cambió, no le prestamos atención. Al cesar, ha vulnerado nuestra predicción y atraído nuestra atención. Veamos un ejemplo histórico. Justo después de que en la ciudad de Nueva York dejaran de funcionar los trenes elevados, la gente llamaba a la policía en mitad de la noche para declarar que algo la había despertado. Solían efectuar la llamada alrededor del momento en que los trenes acostumbraban a pasar por sus apartamentos.

Nos gusta afirmar que ver es creer. No obstante, vemos lo que esperamos ver con tanta frecuencia como vemos lo que en realidad vemos. Uno de los ejemplos más fascinantes al respecto tiene que ver con lo que los investigadores denominan completar. Puede que ya esté enterado de que tenemos un pequeño punto ciego en cada ojo, donde el nervio óptico sale de cada retina por un agujero llamado papila óptica. En esa zona no contamos con fotorreceptores, así que estamos permanentemente ciegos en el punto correspondiente del campo visual. Existen dos razones por las que no solemos darnos cuenta de ello, una prosaica y la otra instructiva. La prosaica es que nuestros dos puntos ciegos no suelen solaparse, con lo cual un ojo compensa al otro.

Pero resulta interesante que tampoco percibamos el punto ciego cuando solo tenemos abierto un ojo. Nuestro sistema visual "completa" la información que falta. Cuando cerramos un ojo y miramos una lujosa alfombra turca o los contornos ondulantes de las vetas de la madera en una mesa de cerezo, no vemos un agujero. Tramas enteras en la alfombra, nudos oscuros completos en la veta de la madera desaparecen de la visión de nuestra retina cuando da la casualidad de que coinciden con los puntos ciegos, mas nuestra experiencia es una extensión sin costuras de texturas y colores. Nuestra corteza visual recurre a los recuerdos de patrones similares y efectúa una corriente constante de predicciones que completan cualquier entrada que falte.

Se completa en todas partes de la imagen visual, no solo en los puntos ciegos. Por ejemplo, les muestro una foto de una playa con un tronco arrastrado por la corriente sobre unas rocas. El límite entre las rocas y el tronco resulta claro y evidente. Sin embargo, si ampliamos la imagen, veremos que las rocas y el tronco presentan una textura y color similares donde se encuentran. En la foto ampliada, el borde del tronco no resulta distinguible de las rocas. Si observamos la escena completa, el borde del tronco está claro, pero en realidad lo inferimos del resto de la imagen. Cuando miramos el mundo, percibimos líneas y límites claros que separan los objetos, pero los datos brutos que entran en nuestros ojos son a menudo ruidosos y ambiguos. Nuestra corteza cerebral completa las secciones que faltan o están desordenadas con lo que piensa que debe ir ahí. Y percibimos una imagen inequívoca.

La predicción en la visión es también una función del modo de moverse nuestros ojos. En el capítulo 3 he mencionado las sacudidas oculares. Unas tres veces cada segundo, nuestros ojos se fijan en un punto y luego saltan de repente a otro. En general, no nos damos cuenta de estos movimientos y normalmente no los controlamos de forma consciente. Y cada vez que nuestros ojos se fijan en un nuevo punto, el patrón que entra en nuestro cerebro cambia por completo desde la última fijación. Así pues, tres veces por segundo nuestro cerebro ve algo muy diferente. Las sacudidas oculares no son totalmente aleatorias. Cuando miramos un rostro, nuestros ojos suelen fijarse primero en un ojo, luego en el otro; van de

un lado a otro y se fijan de forma ocasional en la nariz, la boca, las orejas y otros rasgos. Solo percibimos el "rostro", pero los ojos ven un ojo, otro ojo, la nariz, la boca, un ojo, y así sucesivamente. Me doy cuenta de que les parece que no es así. Lo que perciben es una visión continua del mundo, pero los datos brutos que entran en su cabeza están tan entrecortados como si se hubieran grabado con una cámara de vídeo mal manejada.

Imaginemos ahora que nos encontramos a alguien con una nariz adicional donde debía haber un ojo. Nuestros ojos se fijan primero en un ojo y luego, tras una sacudida ocular, en el otro, pero en su lugar vemos una nariz. Sabríamos sin duda que algo iba mal. Para que esto suceda, nuestro cerebro debe tener una expectativa o predicción de lo que está a punto de ver. Cuando predecimos ojo, pero vemos nariz, se incumple la predicción. Así pues, varias veces por segundo, coincidiendo con cada sacudida ocular, nuestro cerebro realiza una predicción sobre lo que verá a continuación. Cuando esta predicción es errónea, se suscita de inmediato nuestra atención. Por ello nos resulta difícil no mirar a las personas con deformidades. Si viera a una persona con dos narices, ¿le costaría no quedarse mirándola fijamente? Por supuesto, si viviéramos con dicha persona, tras un período nos acabaríamos acostumbrando a las dos narices y dejaría de resultarnos insólito.

Piense ahora en usted mismo. ¿Qué predicciones está realizando? Cuando pasa las páginas de este libro, tiene expectativas de que se doblarán un poco y se moverán de formas predecibles que son diferentes del modo de moverse la cubierta. Si está sentado, está prediciendo que la sensación de presión en su cuerpo persistirá; pero si el asiento se humedeciera, comenzara a deslizarse hacia atrás o sufriera cualquier otro cambio inesperado, dejaría de prestar atención al libro y trataría de imaginar qué está sucediendo. Si emplea cierto tiempo en observarse, puede comenzar a entender que su percepción del mundo, su comprensión del mundo, está íntimamente ligada a la predicción. Su cerebro ha hecho un modelo del mundo y lo está cotejando de forma constante con la realidad. Usted sabe dónde se encuentra y qué está haciendo por la validez de ese modelo.

La predicción no se limita a patrones de información sensorial de bajo nivel como ver y escuchar. Hasta ahora he restringido la exposición a tales ejemplos porque son el modo más sencillo de presentar este marco para comprender la inteligencia. Sin embargo, según el principio de Mountcastle, lo que es válido para las áreas sensoriales de nivel bajo también debe serlo para todas las áreas corticales. El cerebro humano es más inteligente que el de otros animales porque puede realizar predicciones sobre tipos de patrones más abstractos y secuencias de patrones temporales más largas. Para predecir lo que dirá mi esposa cuando me vea, debo saber qué ha dicho en el pasado, que hoy es viernes, que la bolsa reciclable tiene que ponerse en el bordillo los viernes por la noche, que yo no lo hice a tiempo la semana pasada y que su rostro tiene cierta apariencia. Cuando abre la boca, tengo



una predicción bastante sólida de lo que va a decir. En este caso, no sé cuáles serán las palabras exactas, pero sí que me recordará que saque la bolsa. El aspecto importante es que la inteligencia superior no es un tipo diferente de proceso de la inteligencia perceptiva. Se basa en lo fundamental en la misma memoria de la corteza cerebral y algoritmo de predicción.

Hemos de advertir que nuestros tests de inteligencia son en esencia pruebas de predicción. De los juegos de la guardería a los tests de Mensa, las pruebas de coeficiente intelectual se basan en realizar predicciones. Dada una secuencia de números, ¿cuál debe ser el siguiente? Dadas tres visiones diferentes de un objeto con una forma compleja, ¿cuál de las siguientes corresponde también al objeto? ¿La palabra A es a la palabra B lo que la palabra C es a qué palabra?

La ciencia es en sí misma un ejercicio de predicción. Adelantamos nuestro conocimiento del mundo mediante un proceso de hipótesis y verificación. Este libro es en esencia una predicción sobre qué es la inteligencia y cómo funciona el cerebro. Incluso el diseño de productos es sobre todo un proceso predictivo. Al diseñar ropa o automóviles, los ingenieros tratan de predecir qué harán los competidores, qué desearán los consumidores, cuánto costará un nuevo diseño y qué modas tendrán demanda.

La inteligencia se mide por la capacidad de recordar y predecir patrones del mundo, incluidos lenguaje, matemática, propiedades físicas de los objetos y situaciones sociales. Nuestro cerebro percibe patrones del mundo exterior, los almacena como memoria y realiza predicciones basadas en las comparaciones entre lo que ha visto antes y lo que sucede ahora.

\* \* \*

**En** este punto tal vez estén pensando: "Acepto que puedo ser inteligente limitándome a yacer en la oscuridad. Como ha señalado, no necesito actuar para comprender o ser inteligente. ¿Pero no son esas situaciones la excepción? ¿Sostiene realmente que la comprensión y la conducta inteligentes están completamente separadas? En definitiva, ¿no es la conducta, y no la predicción, la que nos hace inteligentes? Después de todo, la conducta es el determinante supremo de la supervivencia".

Es una pregunta razonable y, por supuesto, al final la conducta es lo que importa más para la supervivencia de un animal. Predicción y conducta no están separadas por completo, pero su relación es sutil. En primer lugar, la corteza cerebral apareció en el escenario evolutivo una vez que los animales ya habían desarrollado conductas sofisticadas. Por lo tanto, su valor de supervivencia debe entenderse ante todo atendiendo a las diversas mejoras que podía conferir a las conductas animales existentes. Primero llegó la conducta; luego, la inteligencia. En segundo lugar, menos en el caso del oído, la mayoría de lo que sentimos depende en buena

medida de nuestro modo de movernos en el mundo. Así pues, la predicción y la conducta están estrechamente relacionadas. Exploremos estos temas.

Los mamíferos desarrollaron una gran corteza cerebral porque les proporcionaba cierta ventaja de supervivencia, y esta ventaja acabó implantándose en la conducta. Pero al comienzo la corteza cerebral servía para hacer un uso más efectivo de las conductas existentes, no para crear conductas nuevas por completo. Para aclarar este punto es preciso observar cómo evolucionaron nuestros cerebros.

Surgieron sistemas nerviosos sencillos no mucho después de que las criaturas multicelulares comenzaron a pulular por toda la Tierra, hace cientos de millones de años; pero la historia de la inteligencia real empieza en fecha más reciente con nuestros antepasados reptiles, que lograron conquistar la Tierra. Se esparcieron por todos los continentes y se diversificaron en numerosas especies. Poseían sentidos agudos y cerebros bien desarrollados que les dotaban de una conducta compleja. Sus descendientes directos, los reptiles que sobreviven en la actualidad, siguen conservándolos. Un caimán, por ejemplo, goza de sentidos sofisticados como los suyos o los míos. Posee ojos, oídos, nariz, boca y piel bien desarrollados. Realiza conductas complejas, entre las que se incluyen su capacidad para nadar, correr, ocultarse, cazar, tender emboscadas, tomar el sol, anidar y emparejarse.

¿Qué diferencia hay entre un cerebro humano y uno de reptil? Mucha y poca. Digo poca porque, en líneas generales, todo lo que hay en un cerebro de reptil existe en uno humano. Digo mucha porque un cerebro humano posee algo importantísimo que un reptil no tiene: una gran corteza cerebral. A veces escuchamos a la gente referirse al cerebro "viejo" o cerebro "primitivo". Todo ser humano cuenta con estas estructuras más antiguas en el cerebro, igual que un reptil. Regulan la presión sanguínea, el hambre, el sexo, las emociones y muchos aspectos del movimiento. Cuando estamos de pie, nos balanceamos y andamos, por ejemplo, dependemos mucho del cerebro viejo. Si escuchamos un sonido aterrador, nos entra miedo y empezamos a correr, se debe en buena medida a nuestro cerebro viejo. No se necesita más que un cerebro de reptil para hacer muchas cosas interesantes y útiles. Así pues, ¿qué hace la corteza cerebral si no se requiere estrictamente para ver, escuchar y moverse?

Los mamíferos son más inteligentes que los reptiles debido a su corteza cerebral, que apareció hace decenas de millones de años y es exclusiva de ellos. Lo que hace a los humanos más inteligentes que los restantes mamíferos es sobre todo la gran zona de la corteza cerebral, que se extendió de forma espectacular hace solo un par de millones de años. Recordemos que la constitución de la corteza cerebral emplea un único elemento repetido. La lámina de la corteza cerebral humana es del mismo grosor y tiene prácticamente la misma estructura que la de nuestros parientes mamíferos. Cuando la evolución hace algo grande con mucha rapidez, como en el caso de la corteza cerebral humana, lo logra copiando una estructura

existente. Nos volvimos inteligentes añadiendo muchos más elementos a un algoritmo cortical común. Un error habitual consiste en considerar al cerebro humano la cumbre de miles de millones de años de evolución, lo cual puede ser cierto si pensamos en el sistema nervioso completo. Sin embargo, la corteza cerebral es una estructura relativamente nueva y no lleva en vigor el tiempo suficiente para haber pasado por el refinamiento evolutivo de largo plazo.

Este es el núcleo de mi argumento sobre cómo entender la corteza cerebral y por qué la memoria y la predicción son las claves para desvelar el misterio de la inteligencia. Comenzamos con el cerebro de los reptiles sin corteza cerebral. La evolución descubre que si se agrega un sistema de memoria (la corteza cerebral) a la ruta sensorial del cerebro primitivo el animal adquiere la capacidad de predecir el futuro. Imaginemos que el antiguo cerebro de reptil sigue haciendo sus cosas, pero ahora a la vez se introducen patrones sensoriales en la corteza cerebral, que guarda esta información en su memoria. En un tiempo futuro, cuando el animal se encuentra la misma situación u otra similar, la memoria reconoce las entradas y recuerda lo que sucedió en el pasado. La memoria recordada se compara con el flujo de entradas sensoriales; "completa" la entrada presente y predice lo que se verá a continuación. Al comparar la entrada sensorial presente con la memoria recordada, el animal no solo comprende dónde está, sino que puede ver el futuro.

Ahora imaginemos que la corteza cerebral recuerda no solo lo que el animal ha visto, sino también las conductas que el antiguo cerebro realizaba cuando se hallaba en una situación similar. Ni siquiera hemos de suponer que la corteza cerebral conoce la diferencia entre sensaciones y conducta; para la corteza ambos no son más que patrones. Cuando nuestro animal se encuentra en la misma situación u otra parecida, no solo ve el futuro, sino que recuerda qué conductas condujeron a esa visión del futuro. De este modo, la memoria y la predicción permiten a un animal utilizar sus conductas existentes (el cerebro viejo) de forma más inteligente.

Por ejemplo, imagine que es una rata que está aprendiendo a orientarse por un laberinto por primera vez. Excitada por la incertidumbre o el hambre, utilizará las facultades inherentes de su cerebro viejo para explorar el nuevo entorno: escuchar, mirar, olfatear y arrastrarse cerca de los muros. Toda esta información sensorial es utilizada por su cerebro viejo, pero también pasa a la corteza cerebral, donde se almacena. En algún momento futuro se encontrará en el mismo laberinto. Su corteza cerebral reconocerá la entrada presente como algo que ya ha visto y recordará los patrones guardados que representan lo que sucedió en el pasado. En esencia, le permite ver un trecho corto del futuro. Si fuera una rata que habla, diría: "Oh, reconozco este laberinto, y recuerdo esa esquina". Cuando su corteza cerebral se acuerde de lo que pasó en el pasado, se imaginará encontrando el queso que vio la última vez en el laberinto y cómo llegó hasta él. "Si giro aquí, sé lo que sucederá después. Hay un trozo de queso al final de este

pasillo. Lo veo en mi imaginación.” Mientras corretea por el laberinto, depende de estructuras más antiguas y primitivas para efectuar movimientos como levantar las patas y acicalarse los bigotes. Con su corteza cerebral (relativamente) grande es capaz de recordar los lugares en los que ha estado, volverlos a reconocer en el futuro y efectuar predicciones sobre lo que va a suceder a continuación. Una lagartija sin corteza cerebral posee una capacidad mucho menor para recordar el pasado y tal vez tenga que orientarse de nuevo cada vez en el laberinto. Usted (la rata) comprende el mundo y el futuro inmediato debido a su memoria cortical. Ve imágenes gráficas de las recompensas y peligros que hay más allá de cada decisión y, por lo tanto, se mueve con mayor efectividad por su mundo. Puede ver, literalmente, el futuro.

Pero tenga en cuenta que no está realizando conductas que sean demasiado complejas ni nuevas. No está construyendo un ala delta y volando hacia el queso del final del pasillo. Su corteza cerebral está formando predicciones sobre los patrones sensoriales que le permiten ver el futuro, pero su abanico de conductas disponibles no ha sufrido variaciones. Su facultad para corretear, trepar y explorar sigue pareciéndose mucho a la de una lagartija.

Cuando la corteza cerebral se fue agrandando a lo largo del período evolutivo, fue capaz de recordar cada vez más sobre el mundo. Pudo formar más memorias y realizar más predicciones. La complejidad de dichas memorias y predicciones también aumentó. Pero sucedió algo más notable que condujo a la capacidad única de los seres humanos de tener una conducta inteligente.

La conducta humana trasciende el antiguo repertorio básico de desplazarse alrededor con habilidades de rata. Hemos llevado la evolución de la corteza cerebral a un plano nuevo. Solo los seres humanos crean lenguaje escrito y hablado. Solo los humanos cocinan sus alimentos, cosen ropa, vuelan en avión y construyen rascacielos. Nuestras facultades motoras y planificadoras exceden con creces las de nuestros parientes animales más cercanos. ¿Cómo puede la corteza cerebral, que fue diseñada para realizar predicciones sensoriales, generar la conducta increíblemente compleja única de los humanos? ¿Y cómo pudo evolucionar tan de repente esta conducta superior? Hay dos respuestas para estas preguntas. Una es que el algoritmo cortical es tan potente y flexible que con un pequeño reajuste, exclusivo de los humanos, es capaz de crear conductas nuevas y sofisticadas. La otra respuesta es que la conducta y la predicción son dos caras de la misma cosa. Aunque la corteza cerebral puede imaginar el futuro, solo logra realizar predicciones sensoriales precisas si sabe qué conductas se están poniendo en práctica.

En el sencillo ejemplo de la rata que busca el queso, esta recuerda el laberinto y utiliza la memoria para predecir que verá el queso al torcer la esquina. Pero la rata podría girar a la izquierda o a la derecha; solo si recuerda a la vez el queso y la conducta acertada, “girar a la derecha en la bifurcación”, es capaz de lograr que la

predicción sobre el queso se haga realidad. Aunque es un ejemplo trivial, sirve para poner de manifiesto la íntima relación existente entre la predicción sensorial y la conducta. Toda conducta cambia lo que vemos, escuchamos y sentimos. La mayor parte de lo que sentimos en cualquier momento depende en gran medida de nuestras propias acciones. Mueva el brazo ante su cara. Para predecir que verá su brazo, su corteza cerebral tiene que saber que ha ordenado moverse al brazo. Si la corteza cerebral viera moverse su brazo sin la correspondiente orden motora, usted se sorprendería. El modo más sencillo de interpretarlo sería suponer que su cerebro mueve primero el brazo y luego predice lo que verá. Pero que no es así: la corteza cerebral predice que verá el brazo, y esta predicción es la que origina que la orden motora haga realidad la predicción. Usted primero piensa, lo cual provoca que actúe para lograr que sus pensamientos se hagan realidad.

Veamos ahora los cambios que condujeron a que los seres humanos tengan un repertorio de conductas muy ampliado. ¿Existen diferencias físicas entre la corteza cerebral de un mono y la de un humano que puedan explicar por qué solo el último disfruta del lenguaje y otras conductas complejas? El cerebro humano es unas tres veces mayor que el del chimpancé. Pero no es solo cuestión de tamaño. Una clave para entender el salto en la conducta humana se encuentra en las conexiones entre regiones de la corteza cerebral y partes del cerebro viejo. Para expresarlo con la mayor sencillez, nuestro cerebro está conectado de forma diferente.

Observémoslo con más detenimiento. Todos conocemos los hemisferios izquierdo y derecho del cerebro. Pero existe otra división menos conocida, que es la que debemos analizar para buscar las diferencias humanas. Todos los cerebros, sobre todo los grandes, dividen la corteza cerebral en una parte delantera y otra trasera. Los científicos emplean las palabras *anterior* para la mitad frontal y *posterior* para la de atrás. Un gran surco, llamado cisura de Rolando, separa ambas mitades. La parte posterior de la corteza cerebral contiene las secciones a las que llegan las entradas de los ojos, los oídos y el tacto. Es donde ocurre en buena medida la percepción sensorial. La parte frontal contiene regiones de la corteza cerebral que participan en la planificación y el pensamiento superiores. También comprende la corteza motora, la sección del cerebro más responsable del movimiento de los músculos y, por lo tanto, de crear la conducta.

A medida que la corteza cerebral de los primates se fue agrandando con el paso del tiempo, la mitad anterior alcanzó un tamaño desproporcionado, sobre todo en los humanos. Comparados con los restantes primates y los primeros homínidos, tenemos frentes enormes diseñadas para contener nuestra grandísima corteza cerebral anterior. Pero este agrandamiento no basta para explicar la mejora de nuestra capacidad motora comparada con la de otras criaturas. Nuestra habilidad para efectuar movimientos excepcionalmente complejos proviene del hecho de que nuestra corteza motora efectúa muchas más conexiones con los músculos de

nuestros cuerpos. En otros mamíferos la corteza anterior desempeña un papel menos directo en la conducta motora. La mayoría de los animales dependen en buena medida de las partes más antiguas del cerebro para generar su conducta. En contraste, la corteza cerebral humana usurpó la mayor parte del control motor al resto del cerebro. Si se daña la corteza motora de una rata, puede que no presente carencias apreciables. Si se daña la corteza motora de un humano, se queda parálítico.

La gente suele preguntarme por los delfines. ¿No tienen cerebros enormes? Sí; el delfín posee una corteza cerebral grande. Su estructura es más sencilla (tres capas, frente a nuestras seis) que la humana, pero en todos los restantes parámetros es grande. Es probable que el delfín sea capaz de recordar y entender muchas cosas. Puede reconocer a otros delfines particulares. Es posible que conozca cada rincón del océano donde haya estado. Pero aunque muestran una conducta algo sofisticada, los delfines no se acercan a la nuestra, por lo que cabe conjeturar que su corteza cerebral tiene una influencia menos dominante sobre ella. La cuestión es que la corteza cerebral evolucionó primordialmente para proporcionar una memoria del mundo. Un animal con una corteza cerebral grande podría percibir el mundo como usted y yo. Pero los seres humanos somos únicos en cuanto al papel dominante y avanzado que desempeña esta en nuestra conducta. Por ello tenemos un lenguaje complejo y herramientas refinadas, mientras que otros animales no; por ello podemos escribir novelas, navegar por Internet, enviar sondas a Marte y reflexionar sobre la naturaleza de la energía oscura.

Ahora podemos ver el cuadro completo. La Naturaleza creó primero animales como los reptiles con sentidos sofisticados y conductas complejas, pero relativamente rígidas. Luego descubrió que añadiendo un sistema de memoria e introduciéndole el flujo sensorial el animal podía recordar experiencias pasadas. Cuando el animal se encontraba en una situación igual o parecida, la memoria se recordaba y se establecía una predicción de lo que era probable que sucediera a continuación. De este modo, la inteligencia y el entendimiento comenzaron como un sistema de memoria que introducía predicciones en el flujo sensorial. Estas predicciones son la esencia del entendimiento. Conocer algo significa que puedes realizar predicciones al respecto.

La corteza cerebral evolucionó en dos direcciones. Primero se hizo mayor y más compleja en los tipos de memoria que podía almacenar; era capaz de recordar más cosas y realizar predicciones basadas en relaciones más complicadas. En segundo lugar, empezó a interactuar con el sistema motor del cerebro viejo. Para predecir lo que se iba a escuchar, ver y sentir a continuación necesitaba saber qué acciones se estaban llevando a cabo. En el caso de los humanos, la corteza cerebral ha asumido la mayor parte de nuestra conducta motora. En lugar de limitarse a

realizar predicciones basadas en la conducta del cerebro viejo, la corteza cerebral humana dirige la conducta para satisfacer sus predicciones.

La corteza cerebral humana es particularmente grande y, por lo tanto, tiene una capacidad ingente de memoria. Está prediciendo de modo constante lo que verás, oirás y sentirás, aunque las más de las veces no tenemos conciencia de ello. Estas predicciones son nuestros pensamientos y, cuando se combinan con las entradas sensoriales, nuestras percepciones. A esta visión del cerebro la denomino el marco de memoria-predicción de la inteligencia.

Si la habitación china de Searle contuviera un sistema de memoria similar que pudiera realizar predicciones sobre qué caracteres chinos aparecerían a continuación y qué pasaría después en el relato, estaríamos en situación de garantizar que la habitación entendía chino y comprendía el relato. Ahora podemos ver que Alan Turing estaba equivocado. La predicción, y no la conducta, es la prueba de la inteligencia.

Ya estamos preparados para ahondar en los detalles de esta nueva idea sobre el marco de memoria-predicción del cerebro. Para realizar predicciones sobre acontecimientos futuros, nuestra corteza cerebral tiene que almacenar secuencias de patrones. Para recordar las memorias apropiadas tiene que recuperar patrones por su similitud a patrones pasados (recuerdo autoasociativo). Y, por último, las memorias han de guardarse en una forma invariable a fin de que el conocimiento de acontecimientos pasados sea aplicable a nuevas situaciones que son similares pero no idénticas al pasado. Cómo realiza estas tareas la corteza cerebral, además de la exploración más completa de su jerarquía, constituyen el tema del capítulo siguiente.

## Cómo Funciona la Corteza Cerebral

**Tratar** de imaginar cómo funciona el cerebro es igual que resolver un rompecabezas gigantesco. Se puede abordar de dos modos diferentes. Si se utiliza el planteamiento de "arriba-abajo", se comienza con la imagen que debe presentar el rompecabezas resuelto, empleándola para decidir de qué piezas se prescinde y cuáles hay que buscar. En otro planteamiento, de "abajo-arriba", el centro de atención lo constituye cada una de las piezas. Se estudian sus rasgos inusuales y se busca cómo encajarlas con otras piezas. Si no se cuenta con una ilustración de la solución del rompecabezas, el método de "abajo-arriba" es a veces el único modo de proceder.

El rompecabezas de "comprender el cerebro" es particularmente intimidante. Al carecer de un buen marco para entender la inteligencia, los científicos se han visto obligados a recurrir al planteamiento de "abajo-arriba". Pero la tarea es hercúlea, cuando no imposible, con un rompecabezas tan complejo como el cerebro. Para hacerse una idea de la dificultad, imagine un rompecabezas con varios miles de piezas. Muchas pueden interpretarse de múltiples modos, como si cada una tuviera una imagen en ambas caras pero solo una fuera la acertada. Las piezas apenas presentan formas, así que no se puede estar seguro si dos encajan o no. Muchas de ellas no se utilizarán en la solución definitiva, pero no se sabe cuáles o cuántas. Cada mes llegan nuevas piezas por correo. Algunas reemplazan a las antiguas, como si el creador del rompecabezas dijera: "Sé que lleva trabajando algunos días con estas piezas viejas, pero han resultado estar equivocadas. Lo siento. Utilice estas nuevas en su lugar hasta futuro aviso". Por desgracia, no se tiene idea de cuál será el resultado final; y lo que es peor, tal vez se cuente con algunas ideas, mas son erróneas.

Esta analogía del rompecabezas es una buena descripción de la dificultad que arrostramos al crear una nueva teoría de la corteza cerebral y la inteligencia. Las piezas del rompecabezas son los datos biológicos y conductuales que los científicos han reunido durante más de cien años. Cada mes se publican nuevos artículos, que crean piezas de rompecabezas adicionales. A veces los datos de un científico contradicen los de otro. Como dichos datos pueden interpretarse de diferentes maneras, existe desacuerdo sobre casi todo. Sin una estructura de arriba-abajo no hay consenso sobre lo que buscamos, qué es más importante o cómo interpretar las montañas de información que se han acumulado. Nuestro entendimiento del cerebro se ha quedado atascado en el planteamiento de abajo-arriba. Lo que necesitamos es un marco de arriba-abajo.

El modelo de memoria-predicción puede desempeñar ese papel. Nos puede mostrar cómo comenzar a casar las piezas del rompecabezas. Para realizar predicciones, la corteza cerebral necesita un modo de memorizar y guardar el



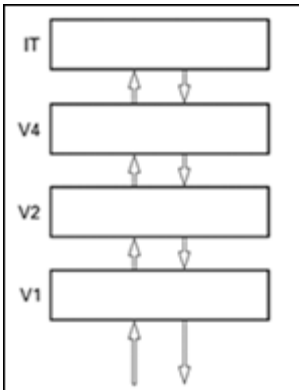
conocimiento sobre secuencias de acontecimientos. Para efectuar predicciones sobre acontecimientos nuevos, la corteza cerebral debe formar representaciones invariables. Nuestro cerebro necesita crear y almacenar un modelo del mundo tal como es, independiente de cómo lo vemos en circunstancias variables. Saber lo que la corteza debe hacer nos guía para comprender su arquitectura, sobre todo su diseño jerárquico y su forma de seis capas.

A medida que vayamos explorando este nuevo marco, presentado en este libro por primera vez, entraré en un nivel de detalle que puede resultar un reto para algunos lectores. Muchos de los conceptos que están a punto de encontrarse son poco conocidos incluso para los expertos en neurociencia. Pero con un poco de esfuerzo creo que cualquiera puede entender los fundamentos de esta nueva estructura. Los capítulos 7 y 8 son mucho menos técnicos y exploran las repercusiones más amplias de la teoría.

Nuestro viaje para resolver el rompecabezas puede pasar ahora a buscar los detalles biológicos que respaldan la hipótesis de la memoria-predicción; sería como lograr separar un gran porcentaje de las piezas del rompecabezas, sabiendo que las relativamente pocas restantes van a revelar la solución definitiva. Una vez que sabemos qué buscar, la tarea se vuelve manejable.

Al mismo tiempo, quiero subrayar que este nuevo marco está incompleto. Hay muchas cosas que todavía no comprendemos, pero hay muchas otras que sí, basándonos en el razonamiento deductivo, experimentos realizados en muchos laboratorios diferentes, y la anatomía conocida. En los últimos cinco o diez años, los investigadores de muchas subespecialidades de la neurociencia han venido explorando ideas similares a la mía, si bien emplean una terminología diferente y no han tratado, hasta donde sé, de colocar estas ideas en una estructura abarcadora. Sí que hablan de procesamiento de arriba-abajo y abajo-arriba, de cómo los patrones se propagan por las regiones sensoriales del cerebro y cómo las representaciones invariables pueden ser importantes. Por ejemplo, los neurocientíficos de Caltech Gabriel Kreiman y Christof Koch, junto con el neurocirujano Itzhak Fried, han descubierto células que se estimulan siempre que una persona ve una foto de Bill Clinton. Una de mis metas es explicar cómo esas células del ex presidente cobran vida. Por supuesto, todas las teorías necesitan hacer predicciones que deben probarse en el laboratorio. Sugiero varias de dichas predicciones en el apéndice. Ahora que sabemos lo que tenemos que buscar, este sistema no parecerá tan complejo nunca más.

En los apartados siguientes de este capítulo iremos ahondando en el funcionamiento del modelo de memoria-predicción. Comenzaremos con la estructura y función a gran escala de la corteza cerebral y avanzaremos en la comprensión de las piezas menores y cómo encajan en el cuadro general.



**Figura 1. Las cuatro primeras regiones visuales en el reconocimiento de objetos.**

### **Representaciones Invariables**

Ya he presentado la imagen de la corteza cerebral como una lámina de células del tamaño de una servilleta grande y el grosor de seis tarjetas de visita, en la cual las conexiones entre las diversas regiones otorgan al todo una estructura jerárquica. Ahora quiero dibujar otro cuadro de la corteza que destaque su conectividad jerárquica. Imaginemos que cortamos la servilleta en regiones funcionales diferentes —secciones de la corteza cerebral que se especializan en ciertas tareas— y amontonamos dichas regiones unas encima de otras como si fueran tortitas. Si cortamos ese montón y lo vemos desde un lado, obtenemos la figura 1. Ojo, la corteza cerebral no tiene esa apariencia en realidad, pero la imagen nos ayudará a visualizar cómo fluye la información. He mostrado cuatro regiones corticales en las que la entrada sensorial penetra por abajo, la región inferior, y fluye hacia arriba de una región a otra. Hemos de tener en cuenta que la información fluye en ambos sentidos.

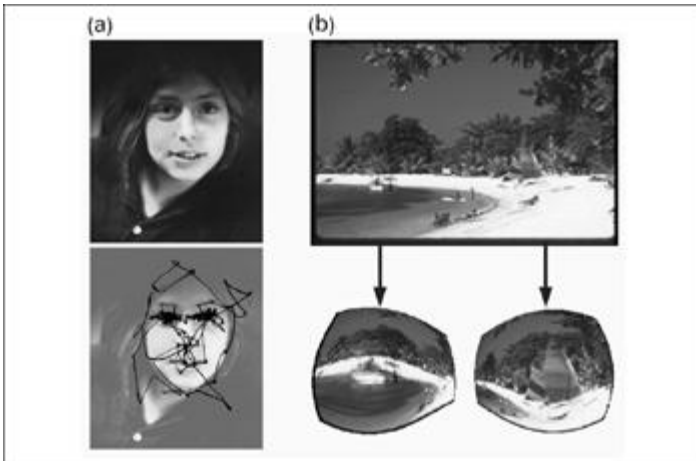
La figura 1 representa las cuatro primeras regiones visuales que participan en el reconocimiento de objetos: cómo se logra ver y reconocer un gato, una catedral, a tu madre, la Gran Muralla china, tu nombre. Los biólogos las denominan V1, V2, V4 e IT. La entrada visual representada por la flecha inferior de la figura se origina en las retinas de ambos ojos y es transmitida a V1. Esta entrada puede concebirse como los patrones en cambio constante transportados por un millón aproximado de axones que se unen para formar el nervio óptico.

Ya hemos hablado de los patrones espaciales y temporales, pero merece la pena refrescar la memoria porque me estaré refiriendo a ellos con frecuencia. Recordemos que la corteza cerebral es una gran lámina de tejido que contiene áreas funcionales especializadas en determinadas tareas. Estas regiones están conectadas mediante grandes manojos de axones o fibras que transfieren información de una región a otra toda a la vez. En cualquier momento, un conjunto de fibras estimula impulsos eléctricos, llamados potenciales de acción,

mientras que otros permanecen quietos. La actividad colectiva en un manajo de fibras es lo que se entiende por *patrón*. El patrón que llega a V1 puede ser espacial, como cuando nuestros ojos se posan por un instante sobre un objeto, y temporal, como cuando nuestros ojos se mueven por el objeto.

Como ya he señalado, unas tres veces por segundo nuestros ojos efectúan un movimiento rápido, llamado sacudida ocular, y una detención, llamada fijación. Si un científico le equipara con un aparato que siguiera los movimientos oculares, le sorprendería descubrir lo entrecortadas que son sus sacudidas oculares, dado que experimenta su visión como si fuera continua y estable. La figura 2a muestra cómo se movían los ojos de una persona mientras miraba una cara. Se advierte que las fijaciones no son aleatorias. Ahora imaginemos que podemos ver el patrón de actividad que llega a V1 desde los ojos de esa persona. Cambia por completo con cada sacudida ocular. Varias veces por segundo la corteza cerebral visual ve un patrón completamente nuevo.

Cabría pensar: "Bien, pero sigue siendo la misma cara, solo que cambiada". Es verdad hasta cierto punto, mas no tanto como piensa. Los receptores de luz de nuestra retina están distribuidos de forma irregular. Se concentran densamente en el centro, llamado fovea, y se van espaciando poco a poco en la periferia. El resultado es que la imagen retinal transmitida al área visual primaria, V1, está muy distorsionada. Cuando nuestros ojos se fijan en la nariz de una cara frente a un ojo de la misma cara, la entrada visual es muy diferente, como si se estuviera contemplando a través de un objetivo de ojo de pez distorsionante que cambia bruscamente de un lado a otro. No obstante, cuando vemos la cara, no parece distorsionada, ni tampoco parece que esté dando saltos. La mayor parte del tiempo ni siquiera tenemos conciencia de que el patrón de la retina ha cambiado, y ya no digamos de forma tan marcada. Nos limitamos a ver una "cara". (La figura 2b muestra este efecto en la vista de un paisaje de playa.) Es una repetición del misterio de la representación invariable del que hablamos en el capítulo 4 sobre la memoria. Lo que "percibimos" no es lo que V1 ve. ¿Cómo sabe nuestro cerebro que está mirando la misma cara y por qué no nos damos cuenta de que las entradas van cambiando y están distorsionadas?



**Figura 2a. Cómo efectúa sacudidas oculares el ojo por una cara humana. Figura 2b. Distorsión causada por la distribución irregular de receptores en la retina.**

Si colocamos una sonda en V1 y observamos cómo responden las células individuales, descubrimos que una célula determinada solo se estimula en respuesta a la entrada visual de una diminuta parte de nuestra retina. Este experimento se ha realizado muchas veces y es uno de los soportes principales de la investigación sobre la visión. Cada neurona de V1 tiene un denominado campo receptivo que es muy específico para una parte diminuta de nuestro campo de visión total, es decir, el mundo completo que hay ante nuestros ojos. Las células de V1 parecen carecer de todo conocimiento sobre las caras, coches, libros u otros objetos significativos que vemos continuamente; no "conocen" más que una diminuta porción, del tamaño de la punta de un alfiler, del mundo visual.

Cada célula de V1 también está adaptada para tipos específicos de patrones de entrada. Por ejemplo, una célula particular podría estimularse mucho cuando ve una línea o un borde inclinado a treinta grados dentro de su campo receptivo. Ese borde tiene poco significado en sí mismo; podría formar parte de cualquier objeto, una tabla del suelo, el tronco de una palmera distante, el lateral de una letra *M*, o cualesquiera de otras posibilidades aparentemente infinitas. Con cada nueva fijación, el campo receptivo de la célula va a detenerse en una porción nueva y completamente diferente del espacio visual. En algunas fijaciones la célula se estimulará; en otras lo hará apenas o en absoluto. Así pues, cada vez que realizamos una sacudida ocular, es posible que muchas células de V1 cambien de actividad.

Sin embargo, algo mágico sucede si se coloca una sonda en la región superior que se muestra en la figura 1, la región IT. Ahí encontramos algunas células que se activan y permanecen estimuladas cuando aparecen objetos completos en el campo visual. Por ejemplo, podríamos encontrar una célula que se activa con fuerza siempre que es visible una cara. Esta célula permanece activa durante el

tiempo en que nuestros ojos miran una cara en cualquier lugar de nuestro campo de visión. No se activa y desactiva con cada sacudida ocular, como las células de V1. El campo receptivo de esta célula de IT cubre el total del espacio visual y entra en actividad cuando ve caras.

Reconstruyamos el misterio. En el curso de abarcar cuatro estadios corticales de la retina a IT, las células han pasado de dedicarse al reconocimiento de rasgos diminutos que cambian de prisa y presentan una especificidad espacial a especializarse en el reconocimiento de objetos, estimulándose de forma continua y careciendo de especificidad espacial. La célula IT nos indica que estamos viendo una cara en cualquier lugar de nuestro campo de visión. Esta célula, comúnmente llamada célula de cara, se estimulará prescindiendo de que dicha cara esté ladeada, rotada o algo tapada. Forma parte de una representación invariable de "cara".

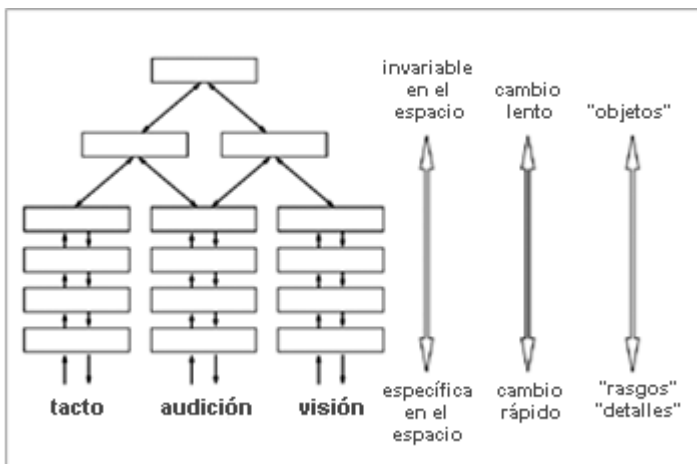
Estas palabras escritas hacen que parezca muy sencillo. Cuatro rápidas etapas y ya está, reconocemos una cara. Ningún programa informático o fórmula matemática ha resuelto este problema con nada que se aproxime a la solidez y generalidad de un cerebro humano. No obstante, sabemos que este lo resuelve en unos cuantos pasos, así que la respuesta no puede ser tan difícil. Una de las metas primordiales de este capítulo es explicar cómo surge una célula de cara, sea de Bill Clinton u otra. Llegaremos a ello, pero antes debemos atender muchos otros aspectos.

Echemos otra ojeada a la figura 1. Podemos ver que la información también fluye de las regiones inferiores a las superiores por una red de conexiones de realimentación. Se trata de manojos de axones que van de las regiones superiores como IT a regiones inferiores como V4, V2 y V1. Además, en la corteza cerebral visual existen las mismas conexiones de realimentación —si no más— que de alimentación.

Durante muchos años, la mayoría de los científicos han ignorado estas conexiones de realimentación. Si su entendimiento del cerebro se centraba en cómo la corteza cerebral aceptaba las entradas, las procesaba y luego actuaba en consecuencia, no necesitaban realimentación. No se precisaban más que conexiones de alimentación que condujeran de las secciones sensoriales a las motoras de la corteza cerebral. Pero cuando comenzamos a darnos cuenta de que la función central de esta es realizar predicciones, tenemos que incorporar la realimentación al modelo; el cerebro ha de enviar información que refluya a la región que recibe primero las entradas. La predicción requiere establecer una comparación entre lo que sucede y lo que se espera que suceda. Lo que sucede en realidad fluye hacia arriba y lo que se espera que suceda lo hace hacia abajo.

El mismo proceso de alimentación y realimentación ocurre en todas las áreas de la corteza cerebral, abarcando todos nuestros sentidos. La figura 3 muestra nuestro montón de tortitas visual junto a montones similares para el oído y el tacto.

También presenta unas cuantas regiones corticales superiores, a veces llamadas áreas de asociación, que reciben e integran entradas de varios sentidos diferentes, como el oído más el tacto y la visión. Mientras que la figura 1 se basa en la conectividad conocida entre cuatro regiones bien estudiadas de la corteza cerebral, la figura 3 no es más que un diagrama conceptual que no pretende captar las regiones corticales reales. En un cerebro humano hay docenas de regiones corticales interconectadas de todo tipo de modos. De hecho, la mayor parte de la corteza cerebral humana está compuesta por áreas de asociación. La caracterización caricaturizada que se muestra en esta y las siguientes figuras pretende ayudarles a comprender lo que sucede sin confundirles en ningún aspecto importante.



**Figura 3. Formación de representaciones invariables en el oído, la visión y el tacto.**

La transformación —de cambio rápido a lento y de específica en el espacio a invariable en el espacio— está bien documentada para la visión. Y aunque existe un volumen de pruebas menor para demostrarlo, muchos neurocientíficos creen que se descubrirá que lo mismo ocurre en todas las áreas sensoriales de la corteza cerebral, no solo en la visión.

Tomemos el oído. Cuando alguien te habla, los cambios en la presión del sonido ocurren con mucha rapidez; los patrones que entran en el área auditiva primaria, llamada A1, cambian igual de rápido. Sin embargo, si lográramos introducir una sonda más arriba en el flujo auditivo, podríamos encontrar células invariables que responden a palabras o incluso, en algunos casos, a expresiones. Tal vez nuestra corteza cerebral auditiva posea un grupo de células que se estimulen cuando escuchamos "gracias" y otro grupo que lo hagan con la expresión "buenos días". Estas células permanecerían activas mientras dure un enunciado, suponiendo que lo reconozcamos.

Los patrones recibidos por la primera área auditiva pueden variar mucho. Una palabra puede decirse con diferentes acentos, en tonos distintos o a velocidades diversas. Pero más arriba, en la corteza cerebral, esos rasgos de bajo nivel no importan; una palabra es una palabra prescindiendo de los detalles acústicos. Lo mismo cabe aplicar a la música. Se puede escuchar *Tres ratones ciegos* tocada al piano, al clarinete o cantada por un niño, y la región A1 recibe un patrón completamente diferente en cada caso. Pero una sonda colocada en una región auditiva superior debería encontrar células que se estimulan de forma constante cada vez que se toca *Tres ratones ciegos* sin que importe el instrumento, compás u otros detalles. Este experimento particular no se ha realizado, por supuesto, ya que es demasiado invasivo para los seres humanos, pero si se acepta que debe haber un algoritmo cortical común, se puede dar por sentado que dichas células existen. Vemos el mismo tipo de realimentación, predicción y recuerdo invariable en la corteza cerebral auditiva que observamos en el sistema visual.

Por último, el tacto tiene que comportarse de la misma manera. De nuevo, no se han realizado los experimentos definitivos, aunque se está investigando con monos en máquinas de imágenes cerebrales de alta resolución. Mientras estoy sentado escribiendo, sostengo una pluma en la mano. Toco el capuchón y mis dedos acarician su gancho de metal. Los patrones que entran en mi corteza cerebral somatosensorial desde los sensores de tacto de mi piel cambian con rapidez a medida que mis dedos se mueven, pero percibo una pluma constante. En un momento dado podría flexionar el gancho de metal con los dedos y al siguiente hacerlo con dedos diferentes o incluso con los labios. Son entradas muy distintas que llegan a emplazamientos diferentes de la corteza cerebral somatosensorial. Sin embargo, nuestra sonda volvería a encontrar células en regiones separadas varios pasos de la entrada primaria que responden de forma invariable a "pluma". Permanecerían activas mientras acariciara la pluma y no les importaría qué dedos o partes de mi cuerpo empleo para tocarla.

Pensemos en esto. Con el oído y el tacto no se puede reconocer un objeto con una entrada momentánea. El patrón proveniente de los oídos o los sensores de tacto de la piel no contiene información suficiente en ningún punto del tiempo para indicar lo que se está escuchando o sintiendo. Cuando percibimos una serie de patrones auditivos como una melodía, una palabra hablada o un portazo, y cuando percibimos un objeto táctil como una pluma, el único modo de hacerlo es emplear el flujo de entrada a lo largo del tiempo. No podemos reconocer una melodía escuchando una nota, ni el tacto de una pluma con un solo toque. Por lo tanto, la actividad neuronal correspondiente a la percepción mental de los objetos, como las palabras habladas, debe durar más en el tiempo que los patrones de entrada individuales. Este no es más que otro modo de llegar a la misma conclusión: cuanto más se suba en la corteza cerebral, deben verse menos cambios a lo largo del tiempo.

La visión también es un flujo de entradas basado en el tiempo y funciona del mismo modo general que el oído y el tacto, pero como tenemos la facultad de reconocer objetos particulares con una sola fijación, confunde el cuadro. En efecto, esta capacidad de reconocer patrones espaciales durante una breve fijación ha desorientado durante muchos años a los investigadores que trabajan en la visión de las máquinas y los animales. En general, han pasado por alto la naturaleza crucial del tiempo. Aunque en condiciones de laboratorio se puede hacer que los seres humanos reconozcan objetos sin mover los ojos, no es la norma. La visión normal, como la suya al leer este libro, requiere un movimiento ocular constante.

## **Integración de los Sentidos**

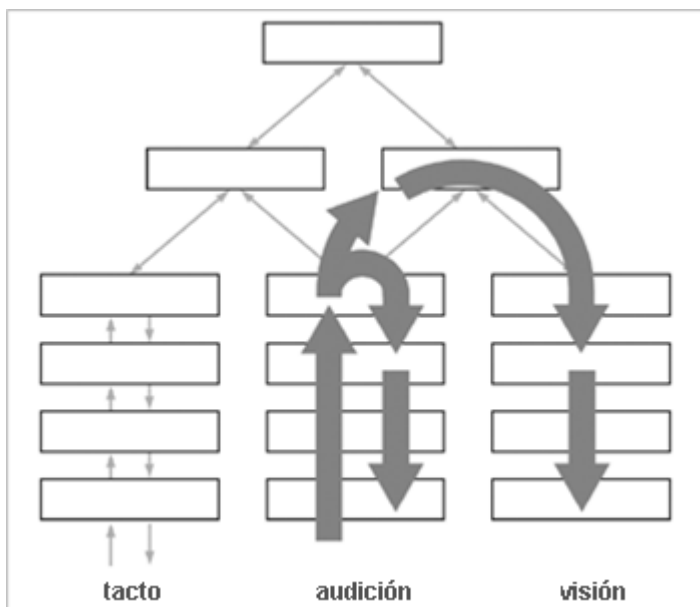
¿Y qué pasa con las áreas de asociación? Hasta ahora hemos visto subir y bajar los flujos de información en un área sensorial particular de la corteza cerebral. El flujo descendente completa la entrada en curso y realiza predicciones sobre lo que se experimentará a continuación. Ocurre el mismo proceso entre los sentidos —es decir, entre la vista, el sonido, el tacto y muchos otros—. Por ejemplo, algo que escucho puede llevar a una predicción de lo que debo ver o sentir. Ahora estoy escribiendo en mi dormitorio. Nuestra gata *Keo* tiene un collar que cascabelea cuando camina. Escucho su cascabeleo aproximarse desde el pasillo. Por esta entrada auditiva reconozco a mi gata; vuelvo la cabeza hacia el pasillo y aparece *Keo*. Espero verla basándome en su sonido. Si no hubiera entrado, o hubiera aparecido otro animal, me habría sorprendido. En este ejemplo, una entrada auditiva creó primero un reconocimiento auditivo de *Keo*. La información fluyó hacia arriba en la jerarquía auditiva hasta un área de asociación que conecta la visión con el oído. Entonces la representación volvió a fluir hacia abajo en las jerarquías auditiva y visual, conduciendo a predicciones auditivas y visuales. La figura 4 lo ilustra.

Este tipo de predicción multisensorial ocurre de continuo. Tuerzo hacia fuera el gancho de mi pluma, siento cómo se suelta de mis dedos y espero escuchar un chasquido cuando golpea el cañón del capuchón. Si no escuchara el chasquido después de soltar el gancho, me sorprendería. Mi cerebro predice con precisión cuándo escucharé el sonido y cómo será. Para que ocurra esta predicción, la información ha fluido hacia arriba por la corteza cerebral somatosensorial y de nuevo hacia abajo por la corteza cerebral somatosensorial y auditiva, llevando a la predicción de que escucharía y sentiría un chasquido.

Otro ejemplo: voy en bicicleta al trabajo varios días a la semana. Esas mañanas entro en el garaje, cojo la bicicleta, le doy la vuelta y la saco rodando hasta la puerta. En el proceso recibo muchas entradas visuales, táctiles y auditivas. La bicicleta golpea la jamba de la puerta, la cadena repiquetea, me doy con un pedal en la pierna y la rueda gira cuando roza el suelo. En el proceso de transportar la bicicleta fuera del garaje, mi cerebro se topa con un aluvión de sensaciones de vista, sonido y tacto. Cada flujo de entrada realiza predicciones para las demás de



un modo tan coordinado que sorprende. Las cosas que veo conducen a predicciones precisas sobre las cosas que siento y escucho, y viceversa. Ver cómo la bicicleta golpea la jamba de la puerta me hace que espere escuchar un sonido particular y sentir su bote hacia arriba. Sentir que el pedal me da en la pierna me incita a mirar hacia abajo y predecir que veré el pedal justo donde lo siento. Las predicciones son tan precisas que me daría cuenta si alguna de estas entradas estuviera levemente descoordinada o no fuera usual. La información fluye de modo simultáneo hacia arriba y abajo de las jerarquías sensoriales para crear una experiencia sensorial unificada que supone predicción en todos los sentidos.



**Figura 4. La información fluye arriba y abajo en las jerarquías sensoriales para formar predicciones y crear una experiencia sensorial unificada.**

Intenten este experimento. Dejen de leer y hagan algo, una actividad que entrañe mover el cuerpo y manipular un objeto. Por ejemplo, vayan al lavabo y abran el grifo. Ahora, mientras lo hacen, traten de percibir todo sonido, sensación de tacto y entradas visuales cambiantes. Tendrán que concentrarse. Cada acción está íntimamente ligada con visiones, sonidos y sensaciones de tacto. Levanten o giren el mando del grifo: su cerebro espera sentir presión sobre la piel y resistencia en los músculos. Ustedes esperan ver y sentir moverse el mando, y ver y escuchar el agua. Cuando el agua golpea el lavabo, esperan escuchar un sonido diferente y ver y sentir la salpicadura.

Todas las pisadas producen un sonido que ustedes siempre anticipan, tengan conciencia de ello o no. Incluso el simple acto de sujetar este libro lleva a muchas predicciones sensoriales. Imaginen que sintieran y escucharan cerrarse el libro, pero visualmente permaneciera abierto. Se sentirían sorprendidos y confusos. Como vimos en el caso de la puerta modificada en el experimento presentado en

el capítulo 5, hacemos predicciones constantes del mundo coordinadas con todos los sentidos. Cuando me concentro en todas las pequeñas sensaciones, me asombra lo plenamente integradas que están nuestras predicciones perceptivas. Aunque estas percepciones pueden parecer sencillas o triviales, no hay que olvidar su carácter omnipresente y que solo pueden ocurrir mediante la coordinación de los patrones que fluyen en ambos sentidos de la jerarquía cortical.

Una vez que se comprende la gran interconexión de los sentidos, se llega a la conclusión de que toda la corteza cerebral, todas las áreas sensoriales y de asociación, actúan como una sola. Sí, tenemos una corteza cerebral visual, pero no es más que un componente de un sistema sensorial único y abarcador: visiones, sonidos, tactos y todo lo demás combinado fluyendo arriba y abajo de una única jerarquía con múltiples ramas.

Una cuestión más: todas las predicciones se aprenden por experiencia. Esperamos que los ganchos de las plumas hagan determinados sonidos en el presente y el futuro porque los han hecho en el pasado. Las bicicletas que se golpean en los garajes se ven, sienten y suenan de modos predecibles. No hemos nacido con ninguno de estos conocimientos; los aprendimos gracias a la grandísima capacidad de nuestra corteza cerebral para recordar patrones. Si hay patrones constantes entre las entradas que fluyen a nuestros cerebros, nuestra corteza cerebral los usará para predecir acontecimientos futuros.

Aunque las figuras 3 y 4 no representan la corteza cerebral motora, pueden imaginarla como otro montón jerárquico de tortitas, semejante al montón sensorial, conectado a los sistemas sensoriales por las áreas de asociación (tal vez con conexiones más íntimas con la corteza cerebral somatosensorial para efectuar movimientos corporales). De este modo, la corteza cerebral motora se comporta casi igual que una región sensorial. Una entrada en un área sensorial puede fluir hacia arriba hasta un área de asociación, lo que puede conducir a que un patrón fluya hacia abajo a la corteza cerebral motora y se obtenga como resultado una conducta. Del mismo modo que una entrada visual puede conducir a que los patrones fluyan hacia abajo a las secciones auditiva y táctil de la corteza cerebral, también puede llevar a que un patrón fluya hacia abajo a la sección motora de la corteza cerebral. En el primer caso, interpretamos esos patrones que fluyen hacia abajo como órdenes motoras. Como señaló Mountcastle, la corteza cerebral motora se parece a la corteza cerebral sensorial. Así pues, su forma de procesar predicciones sensoriales que fluyen hacia abajo es similar a su modo de procesar órdenes motoras que fluyen en el mismo sentido.

Veremos pronto que no existen áreas puramente sensoriales o motoras en la corteza cerebral. Los patrones sensoriales fluyen a todas partes, y luego vuelven a descender por un área de la jerarquía, llevando a predicciones o conductas motoras. Aunque la corteza cerebral motora posee algunos atributos especiales, es razonable concebirla como parte de un gran sistema de memoria-predicción

jerárquico. Es casi como otro sentido. Ver, oír, tocar y actuar están profundamente entrelazados.

## **Una Nueva Visión de V1**

El siguiente paso para desentrañar la arquitectura de la corteza cerebral requiere que miremos todas las regiones corticales de una manera nueva. Sabemos que las regiones superiores de la jerarquía cortical forman representaciones invariables, ¿pero por qué esta importante función solo ocurre en la cima? Teniendo en mente la noción de simetría de Mountcastle, comencé a explorar los modos diferentes en que las regiones corticales podrían estar conectadas.

La figura 1 describe las cuatro regiones clásicas de la ruta visual, V1, V2, V4 e IT, con V1 en el fondo del montón, V2 y V4 encima, e IT como remate. Se las suele considerar y mostrar como una región única y continua. Así pues, se supone que todas las células de V1 hacen cosas similares, aunque con diferentes partes del campo visual. Todas las células de V2 efectúan el mismo tipo de tareas. Todas las células de V4 poseen una especialización semejante.

En este planteamiento tradicional, cuando la imagen de un rostro entra en la región de V1, las células que hay en ella crean un esbozo tosco de él sirviéndose de sencillos segmentos de línea y otros rasgos elementales. El esbozo se pasa a V2, donde la imagen recibe un análisis algo más sofisticado de los rasgos faciales, que a continuación se envía a V4, y así sucesivamente. La invariancia y el reconocimiento del rostro solo se logran cuando la entrada llega a la cima, IT.

Por desgracia, este planteamiento uniforme de las primeras regiones corticales como V1, V2 y V4 presenta algunos problemas. Una vez más, ¿por qué solo deben aparecer representaciones invariables en IT? Si todas las regiones de la corteza cerebral realizan la misma función, ¿por qué IT ha de ser especial?

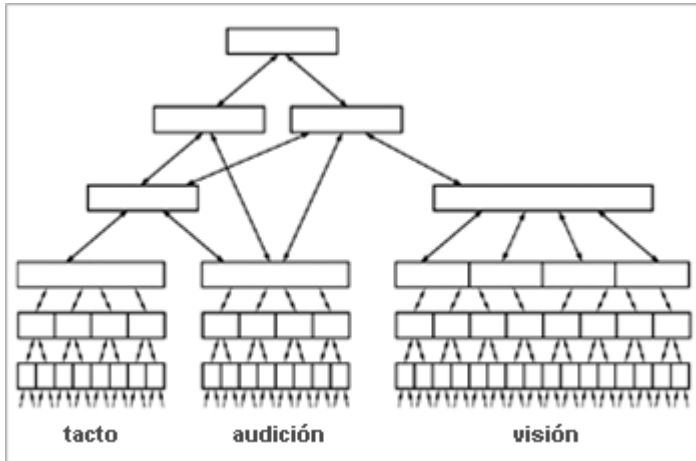
En segundo lugar, el rostro puede aparecer en el lado izquierdo o el derecho de nuestro V1 y lo reconoceríamos. Pero los experimentos muestran con claridad que las partes no contiguas de V1 no están directamente conectadas; el lado izquierdo de V1 no puede saber de forma directa qué está viendo el lado derecho. Retrocedamos para reflexionar sobre ello. Sin duda, las distintas partes de V1 están haciendo algo similar, puesto que todas pueden participar en el reconocimiento de un rostro, mas al mismo tiempo son físicamente independientes. Las subregiones o grupos de V1 están desconectadas desde el punto de vista físico, pero hacen lo mismo.

Por último, los experimentos muestran que todas las regiones superiores de la corteza cerebral reciben entradas convergentes desde dos o más regiones sensoriales de debajo (figura 3). En los cerebros reales una docena de regiones pueden converger en un área de asociación. Pero en las representaciones

tradicionales las regiones sensoriales inferiores como V1, V2 y V4 parecen tener un tipo diferente de conectividad. Parece como si cada una no tuviera más que una única fuente de entradas —solo una flecha fluye hacia arriba desde el fondo— y no hay una convergencia clara de entradas desde las diferentes regiones. V2 obtiene sus entradas de V1, y eso es todo. ¿Por qué algunas regiones corticales reciben entradas convergentes y otras no? Esto tampoco concuerda con la idea de Mountcastle del algoritmo cortical común.

Por esta y otras razones, he llegado a pensar que V1, V2 y V4 no deben considerarse regiones corticales únicas, sino cada cual una reunión de muchas subregiones menores. Volvamos a la analogía de la servilleta grande, que constituye una versión aplanada de toda la corteza cerebral. Pongamos que fuéramos a utilizar una pluma para marcar todas las regiones funcionales de la corteza cerebral en nuestra servilleta cortical. La región mayor con creces es V1, el área visual primaria. A continuación iría V2. Son enormes comparadas con la mayoría de las regiones. Lo que sugiero es que V1 debe considerarse en realidad muchas regiones muy pequeñas. En lugar de una gran área de la servilleta, dibujaríamos muchas áreas pequeñas que juntas ocuparían el área que suele asignarse a V1. En otras palabras, V1 está compuesta por numerosas pequeñas áreas corticales separadas que solo están conectadas con sus vecinas de forma indirecta a través de regiones que se encuentran por encima en la jerarquía. V1 tendría un mayor número de pequeñas subregiones que cualquier área visual. V2 también estaría compuesta por subregiones, aunque menos y algo mayores. Lo mismo cabe aplicar a V4. Pero cuando llegamos a la región superior, IT, nos encontraríamos realmente con una única región, motivo por el cual sus células gozan de una vista aérea de todo el mundo visual.

Existe una grata simetría. Echemos una ojeada a la figura 5, que muestra la misma jerarquía que la figura 3, salvo porque refleja las jerarquías sensoriales como las acabo de describir. Adviértase que ahora la corteza cerebral parece similar en todas partes. escoja una región y encontrará muchas regiones inferiores que proporcionan entradas sensoriales convergentes. La región receptora reenvía proyecciones a sus regiones de entrada, indicándoles qué patrones deben esperar ver a continuación. Las áreas de asociación superiores unen información proveniente de múltiples sentidos como la visión y el tacto. Una región inferior como V2 une la información de subregiones separadas dentro de V1. Una región no sabe —no puede saber— qué significan cualesquiera de esas entradas. Una subregión de V2 no necesita saber que está manejando entradas visuales procedentes de múltiples partes de V1. Un área de asociación no precisa saber que está manejando entradas de la visión y el oído. En su lugar, la labor de cualquier región cortical es descubrir cómo se relacionan sus entradas, memorizar la secuencia de correlaciones entre ellas y utilizar esta memoria para predecir cómo se comportarán las entradas en el futuro. La corteza cerebral es la corteza cerebral. El mismo proceso ocurre en todas partes: un algoritmo cortical común.



**Figura 5. Planteamiento alternativo de la jerarquía cortical**

Esta nueva representación jerárquica nos ayuda a comprender el proceso de creación de representaciones invariables. Observemos con mayor detenimiento cómo funciona en la visión. En el primer nivel de procesamiento, el lado izquierdo del espacio visual es diferente del derecho, del mismo modo que el oído es diferente de la vista. V1 izquierdo y V1 derecho forman el mismo tipo de representaciones debido solo a que son expuestos a patrones similares de la vida. Al igual que el oído y la vista, cabe considerarlos flujos sensoriales separados que se unen en planos superiores.

De igual modo, las regiones menores dentro de V2 y V4 son áreas de asociación de la visión. (Las subregiones pueden solaparse, lo cual no cambiaría en esencia su forma de funcionar.) Interpretar la corteza cerebral visual de esta manera no contradice o cambia nada de lo que conocemos sobre su anatomía. La información fluye hacia arriba y hacia abajo en todas las ramas del árbol de memoria jerárquico. Un patrón del campo visual izquierdo puede llevar a una predicción en el campo visual derecho, del mismo modo que el cascabel de mi gata puede incitar la predicción visual de que está entrando en mi dormitorio.

El resultado más importante de esta nueva representación de la jerarquía cortical es que ahora podemos afirmar que todas y cada una de las regiones corticales forman representaciones invariables. En el planteamiento antiguo no contábamos con representaciones invariables completas —como rostros— hasta que las entradas alcanzaban la capa suprema, IT, que ve el mundo visual completo. Ahora podemos afirmar que las representaciones invariables son ubicuas; se forman en todas las regiones corticales. La invariancia no es algo que aparezca de forma mágica cuando se alcanzan las regiones superiores de la corteza cerebral, como IT. Cada región forma representaciones invariables extraídas del área de entrada inferior desde la perspectiva jerárquica. Así, las subregiones de V4, V2 e IT crean representaciones invariables basadas en lo que fluye a ellas. Puede que solo vean una parte diminuta del mundo y que el vocabulario de objetos sensoriales que

manejen sea más básico, pero realizan la misma labor que IT. Además, las regiones de asociación por encima de IT forman representaciones invariables de patrones procedentes de múltiples sentidos. De este modo, todas las regiones de la corteza cerebral forman representaciones invariables del mundo que tienen por debajo en la jerarquía. Es algo hermoso.

Nuestro rompecabezas ha cambiado. Ya no tenemos que preguntarnos cómo se forman las representaciones invariables en cuatro pasos de abajo arriba, sino cómo se forman en cada una de las regiones corticales, lo cual tiene mucho sentido si nos tomamos en serio la existencia de un algoritmo cortical común. Si una región almacena secuencias de patrones, todas las regiones crean representaciones invariables. El replanteamiento de la jerarquía cortical según las líneas mostradas en la figura 5 posibilita esta interpretación.

## **Un Modelo del Mundo**

¿Por qué la corteza cerebral está construida como una jerarquía?

Podemos pensar sobre el mundo, movernos por él y efectuar predicciones sobre el futuro porque nuestra corteza cerebral ha creado un modelo del mundo. Uno de los conceptos más importantes de este libro es que su estructura jerárquica guarda un modelo de la estructura jerárquica del mundo real. La estructura nido del mundo real se refleja en la estructura nido de nuestra corteza cerebral.

¿Qué entiendo por estructura nido o jerárquica? Pensemos en la música. Las notas se combinan para formar intervalos; los intervalos se combinan para formar frases melódicas; las frases se combinan para formar melodías o canciones; las canciones se combinan en álbumes. Pensemos en el lenguaje escrito. Las letras se combinan para formar sílabas; las sílabas se combinan para formar palabras; las palabras se combinan para formar cláusulas y oraciones. Para considerarlo de un modo totalmente distinto, pensemos en nuestros barrios. Es probable que contengan carreteras, colegios y casas. Las casas tienen habitaciones. Cada habitación tiene paredes, techo, suelo, puerta y una o más ventanas. Cada uno de esos elementos está compuesto de objetos menores. Las ventanas están hechas de cristal, marcos, picaportes y persianas. Los picaportes se componen de piezas menores, como los tornillos.

Dediquemos un momento a observar nuestro entorno. Los patrones de la retina que entran en nuestra corteza cerebral visual primaria se combinan para formar segmentos de línea. Los segmentos de línea se combinan para crear formas más complejas. Estas formas más complejas se combinan para formar objetos como las narices. Las narices se combinan con los ojos y las bocas para formar rostros. Y los rostros se combinan con otras partes del cuerpo para formar la persona que está sentada en la habitación frente a usted.

Todos los objetos de nuestro mundo están compuestos por subobjetos que aparecen siempre juntos; esa es la mera definición de objeto. Cuando asignamos un nombre a algo, lo hacemos porque hay un conjunto de rasgos que van constantemente juntos. Un rostro es un rostro debido a que siempre aparecen juntos dos ojos, una nariz y una boca. Un ojo es un ojo porque siempre aparecen juntos una pupila, un iris, un párpado y demás. Lo mismo cabe afirmar de las sillas, los coches, los árboles, los parques y los países. Y, para finalizar, una canción es una canción porque siempre aparecen juntos en una secuencia una serie de intervalos.

De este modo, el mundo es como una canción. Cada objeto del mundo está compuesto por una reunión de objetos menores, y la mayoría de los objetos forman parte de otros mayores. Esto es lo que entiendo por estructura nido. Una vez que tenemos conciencia de ella, podemos ver una estructura nido en todas partes. De modo análogo, nuestros recuerdos de las cosas y la forma como nuestro cerebro las representa se guardan en la estructura jerárquica de la corteza cerebral. La memoria de nuestra casa no existe en una región de la corteza cerebral. Se guarda en una jerarquía de regiones corticales que refleja la estructura jerárquica de la casa. Las relaciones de gran escala se almacenan en la cima de la jerarquía, y las relaciones de pequeña escala, en el fondo.

El diseño de la corteza cerebral y el método por el cual aprende descubren de forma natural las relaciones jerárquicas del mundo. No nacemos con el conocimiento del lenguaje, las casas o la música. La corteza cerebral posee un algoritmo de aprendizaje inteligente que descubre y capta de forma natural cualquier estructura jerárquica que exista. Cuando falta dicha estructura, caemos en la confusión e incluso en el caos.

En un momento determinado, solo somos capaces de experimentar un subconjunto del mundo. Solo podemos estar en una habitación de nuestra casa, mirando en una dirección. Debido a la jerarquía de la corteza cerebral, logramos saber que estamos en casa, en nuestro salón, mirando una ventana, aunque en ese momento dé la casualidad de que nuestros ojos se hayan fijado en el pasador de la ventana. Las regiones superiores de la corteza cerebral mantienen una representación de nuestra casa, mientras que las inferiores representan las habitaciones, y otras más inferiores miran la ventana. De modo similar, la jerarquía nos permite saber que estamos escuchando una canción y un álbum de música, aunque en un momento determinado solo escuchemos una nota, que por sí misma no nos dice casi nada. Nos permite saber que estamos con nuestra mejor amiga, aunque nuestros ojos se hayan fijado momentáneamente en su mano. Las regiones superiores de la corteza cerebral están atentas al cuadro general, mientras que las áreas inferiores se ocupan de los pequeños detalles que cambian con rapidez.

Puesto que solo podemos tocar, escuchar y ver partes muy pequeñas del mundo en un momento determinado del tiempo, la información que fluye al cerebro llega de forma natural como una secuencia de patrones. La corteza cerebral quiere aprender esas secuencias que aparecen una y otra vez. En algunos casos, como en el de las melodías, una secuencia de patrones llega en un orden rígido, el orden de los intervalos. La mayoría de nosotros conocemos ese tipo de secuencia. Pero voy a utilizar la palabra *secuencia* de un modo más general, más próximo en significado al término matemático *conjunto*. Una secuencia es un conjunto de patrones que suelen acompañarse mutuamente, pero no siempre en un orden fijo. Lo importante es qué patrones de una secuencia se siguen unos a otros en el tiempo, aunque no sea en un orden fijo.

Algunos ejemplos aclararán este punto. Cuando miro su cara, la secuencia de patrones de entrada que veo no es fija, sino que está determinada por mis sacudidas oculares. Una vez podría fijarme en el orden "ojo ojo nariz boca" y un momento después fijarme en el orden "boca ojo nariz ojo". Los componentes de un rostro son una secuencia. Están relacionados estadísticamente y tienden a aparecer juntos en el tiempo, aunque el orden puede variar. Si percibimos "rostro" mientras nos fijamos en "nariz", los patrones probables siguientes serían "ojo" o "boca", pero no "pluma" o "coche".

Cada región de la corteza cerebral ve un flujo de dichos patrones. Si están relacionados de un modo que permite a la región aprender a predecir qué patrón aparecerá a continuación, la región cortical forma una representación persistente o memoria para la secuencia. Aprender secuencias es el requisito básico para formar representaciones invariables de los objetos del mundo real.

La posibilidad de predicción es la definición de la realidad. Si una región de la corteza cerebral descubre que puede moverse con fiabilidad y previsión entre estos patrones de entrada utilizando una serie de movimientos físicos (como las sacudidas oculares de los ojos o las caricias de los dedos) y predecirlos con precisión a medida que se despliegan en el tiempo (como los sonidos que comprende una canción o una palabra hablada), el cerebro interpreta que presentan una relación causal. La probabilidad de que aparezcan numerosos patrones de entrada en la misma relación una y otra vez por pura coincidencia es pequeñísima. Una secuencia predecible de patrones forma parte de un objeto mayor que existe realmente. Así pues, la posibilidad de predicción fiable es un modo seguro de saber que diferentes acontecimientos del mundo están unidos físicamente. Todos los rostros tienen ojos, orejas, boca y nariz. Si el cerebro ve un ojo, efectúa una sacudida ocular y ve otro ojo, efectúa una sacudida ocular más y ve una boca, puede sentirse seguro de que está contemplando un rostro.

Si las regiones corticales pudieran hablar, tal vez dijeran: "Experimento diferentes patrones. A veces no soy capaz de predecir qué patrón veré a continuación. Pero esos conjuntos de patrones están relacionados entre sí. Siempre aparecen juntos y



puedo saltar con confianza entre ellos. Así que siempre que vea uno de esos hechos me referiré a ellos con un nombre común". Es este nombre de grupo el que pasará a las regiones superiores de la corteza cerebral.

Así pues, cabe afirmar que el cerebro almacena secuencias de secuencias. Cada región de la corteza cerebral aprende secuencias, desarrolla lo que denomino "nombres" para las secuencias que conoce y pasa esos nombres a las regiones superiores siguientes de la jerarquía cortical.

## **Secuencias de Secuencias**

A medida que la información avanza hacia arriba desde las regiones sensoriales primarias hasta los niveles superiores, vemos cada vez menos cambios a lo largo del tiempo. En las áreas visuales primarias como V1, el conjunto de células activas cambia con rapidez a medida que entran en la retina nuevos patrones varias veces por segundo. En el área visual IT, los patrones de activación de las células son más estables. ¿Qué sucede ahí? Cada región de la corteza cerebral posee un repertorio de secuencias que conoce, análogo a un repertorio de canciones. Las regiones almacenan esas secuencias semejantes a canciones sobre todas y cada una de las cosas: el sonido de las olas que rompen en la playa, el rostro de tu madre, el camino de tu casa a la tienda de la esquina, la ortografía de la palabra "palomita", o cómo barajar las cartas.

De forma similar a los nombres que tenemos para las canciones, cada región cortical posee un nombre para cada secuencia que conoce. Este "nombre" es un grupo de células cuya activación colectiva representa el conjunto de objetos de la secuencia. (No importa por ahora cómo se selecciona ese grupo de células para representar la secuencia; nos ocuparemos de ello más adelante.) Estas células permanecen activas mientras la secuencia se representa, y es su "nombre" el que se pasa a la región siguiente de la jerarquía. Mientras los patrones de entrada formen parte de una secuencia predecible, la región presenta un "nombre" constante a las regiones superiores siguientes.

Es como si la región estuviera diciendo: "Aquí está el nombre de la secuencia que estoy escuchando, viendo o tocando. No necesitas conocer las notas, bordes o textura individuales. Te haré saber si algo nuevo impredecible sucede". Más en concreto, cabe imaginar a la región IT de la cima de la jerarquía visual transmitiendo a un área de asociación de debajo: "Estoy viendo un rostro. Sí, con cada sacudida ocular los ojos se fijan en partes diferentes del rostro; estoy viendo partes distintas del rostro en sucesión. Pero sigue siendo el mismo rostro. Os lo comunicaré cuando vea algo más". De este modo, una secuencia predecible de acontecimientos queda identificada con un "nombre", un patrón constante de activación celular. Esto ocurre una y otra vez mientras ascendemos en la pirámide jerárquica. Una región podría reconocer una secuencia de sonidos que comprende fonemas (los sonidos que constituyen las palabras) y pasar un patrón que

representa el fonema a la región superior siguiente. Esta reconoce la secuencia de fonemas para crear palabras. La siguiente región superior reconoce secuencias de palabras para crear oraciones, y así sucesivamente. No debe olvidarse que una "secuencia" en las regiones más bajas de la corteza cerebral puede ser muy simple, como un borde visual que se mueve por el espacio.

Al desintegrar secuencias predecibles en "objetos con nombre" en cada región de nuestra jerarquía, logramos cada vez mayor estabilidad a medida que ascendemos, y así se crean las representaciones invariables.

Sucede el efecto contrario cuando un patrón desciende en la jerarquía: los patrones estables se "despliegan" en secuencias. Supongamos que memorizamos el Discurso de Gettysburg cuando estábamos en séptimo grado y ahora queremos recitarlo. En una región de lenguaje superior de nuestra corteza cerebral existe un patrón guardado que representa la famosa alocución de Lincoln. Primero este patrón se despliega en una memoria de la secuencia de las oraciones. En la siguiente región inferior, cada frase se despliega en una memoria de la secuencia de las palabras. En este punto, el patrón desplegado se divide y viaja hacia abajo tanto a la sección auditiva como a la sección motora de la corteza cerebral. Siguiendo la ruta motora, cada palabra se despliega en una secuencia memorizada de fonemas. Y en la región final inferior, cada fonema se despliega en una secuencia de órdenes musculares para crear los sonidos. Cuanto más abajo miremos en la jerarquía, con mayor rapidez cambian los patrones. Un patrón único y constante de la cima de la jerarquía motora acaba llevando a una secuencia compleja y prolongada de sonidos del habla.

La invariancia también funciona a nuestro favor cuando esta información desciende en la jerarquía. Si queremos teclear el Discurso de Gettysburg en lugar de pronunciarlo, comenzamos con el mismo patrón en la cima de la jerarquía. El patrón se despliega en oraciones en la región de debajo. Las oraciones se despliegan en palabras en la región inferior. Hasta ahora no hay diferencia entre pronunciar y teclear el discurso. Pero en el siguiente nivel inferior nuestra corteza cerebral motora toma una ruta diferente. Las palabras se despliegan en letras, y estas, en órdenes musculares a nuestros dedos para que tecleen: "Hace ochenta y siete años, nuestros padres fundaron...". Las memorias de las palabras se manejan como representaciones invariables; no importa si vamos a pronunciarlas, teclearlas o escribirlas a mano. Adviértase que no tenemos que memorizar el discurso dos veces, una para pronunciarlo y otra para escribirlo. Una sola memoria puede adoptar varias formas de conducta. En cualquier región, un patrón invariable puede bifurcarse y seguir una ruta descendente diferente.

En una muestra complementaria de eficiencia, las representaciones de objetos sencillos en el fondo de la jerarquía pueden volver a utilizarse una y otra vez para secuencias diferentes de nivel superior. Por ejemplo, no tenemos que aprender un conjunto de palabras para el Discurso de Gettysburg y otro completamente distinto

para "Tengo un sueño" de Martin Luther King, aunque las dos oraciones contengan algunas palabras iguales. Una jerarquía de secuencias nido permite compartir y reutilizar objetos de nivel inferior; palabras, fonemas y letras no son más que un ejemplo. Es una forma muy eficiente de almacenar información sobre el mundo y su estructura, muy diferente del modo de funcionar de los ordenadores.

El mismo despliegue de secuencias ocurre en las regiones sensoriales y motoras. El proceso nos permite percibir y comprender objetos desde perspectivas diferentes. Si camina hacia su frigorífico para coger un helado, su corteza cerebral visual está activa en múltiples planos. En un nivel superior percibe un "frigorífico" constante. En regiones inferiores esta expectativa visual se rompe en una serie de entradas visuales más localizadas. Ver el frigorífico se compone de fijaciones en el agarrador de la puerta, el dispensador de hielo, los imanes sobre la puerta, un dibujo infantil y demás. En las pocas milésimas de segundo que pasan mientras efectuamos una sacudida ocular de un rasgo a otro del frigorífico, descienden en sucesión las predicciones sobre el resultado de cada sacudida ocular por nuestra jerarquía visual. Mientras dichas predicciones se confirman una sacudida ocular tras otra, nuestras regiones visuales superiores siguen satisfechas con que está viendo en realidad su frigorífico. Advirtamos que en este caso, a diferencia del orden fijo de palabras del Discurso de Gettysburg, la secuencia que ve cuando mira el frigorífico no está fija; el flujo de entradas y los patrones de la memoria recuperada dependen de sus propias acciones. Así pues, en un caso como este, el patrón que se despliega no es una secuencia rígida, pero el resultado final es el mismo: patrones de nivel superior que cambian con lentitud, desplegándose en patrones de nivel inferior que cambian con rapidez.

Nuestra forma de memorizar secuencias y representarlas con un nombre mientras la información asciende y desciende por nuestra jerarquía cortical tal vez les recuerde la jerarquía de mando militar. El general supremo del ejército dice: "Traslade las tropas a Florida para pasar el invierno". Esta sencilla orden de alto nivel se despliega en secuencias cada vez más detalladas de órdenes a medida que va calando en la jerarquía. Los subalternos del general reconocen que la orden requiere una secuencia de pasos, como los preparativos para marcharse, el transporte a Florida y los preparativos de la llegada. Cada uno de estos pasos se descompone en otros más específicos que deben realizar los subordinados. En la base hay miles de soldados rasos efectuando cientos de miles de acciones que dan como resultado el traslado de las tropas. En cada nivel se generan informes de lo que ha sucedido. A medida que van ascendiendo en la jerarquía, se van resumiendo una y otra vez, hasta que en la cima el general recibe una información diaria que dice: "El traslado a Florida va bien". El general no obtiene todos los detalles.

Existe una excepción a esta regla. Si algo va mal y no lo pueden solucionar los subordinados inferiores de la cadena de mando, el asunto va ascendiendo en la

jerarquía hasta que alguien sabe lo que hay que hacer a continuación. El oficial que sabe cómo manejar la situación no la considera una excepción. Lo que para los subordinados constituía un problema imprevisto no es más que la siguiente tarea esperada en una lista. Entonces el oficial emite nuevas órdenes a sus subordinados. La corteza cerebral se comporta de modo similar. Como veremos dentro de poco, cuando ocurren acontecimientos (en otras palabras, patrones) imprevistos, la información al respecto asciende en la jerarquía cortical hasta que alguna región puede manejarla. Si las regiones inferiores no logran predecir qué patrones están viendo, lo consideran un error y lo pasan hacia arriba de la jerarquía. Esto se repite hasta que alguna región prevé el patrón.

\* \* \*

**P**or su diseño, cada región cortical intenta almacenar y recordar secuencias. Pero esta sigue siendo una descripción demasiado simple del cerebro. Necesitamos añadir unas cuantas complejidades más al modelo.

Las entradas de abajo-arriba a una región de la corteza cerebral son patrones de entrada transportados en miles de millones de axones. Estos axones provienen de diferentes regiones y contienen toda clase de patrones. El número de patrones posibles que pueden existir incluso en un millar de axones es mayor que el número de moléculas del Universo. Una región solo verá una diminuta fracción de dichos patrones en su vida.

Por lo tanto, se plantea una pregunta: cuando una sola región almacena secuencias, ¿de qué son? La respuesta es que una región clasifica primero sus entradas como una de un número limitado de posibilidades y luego busca secuencias. Imaginemos que somos una región cortical determinada. Nuestra tarea es clasificar trozos de papel de colores. Nos suministran diez cubos, cada uno etiquetado con una muestra de color. Hay un cubo para el verde, otro para el amarillo, otro más para el rojo, y así sucesivamente. Luego nos entregan trozos de papel de colores, uno por uno, y nos dicen que los clasifiquemos por el color. Cada papel que recibimos es ligeramente diferente. Como existe un número infinito de colores en el mundo, nunca tenemos dos trozos de papel con el mismo color exacto. A veces es fácil indicar en qué cubo debe colocarse el papel coloreado, pero otras resulta difícil. Un papel que es medio rojo y naranja podría ir en cualquiera de esos cubos, pero tenemos que asignarle uno, rojo o naranja, aunque la selección resulte aleatoria. (Lo importante de este ejercicio es mostrar que el cerebro debe clasificar patrones. Las regiones de la corteza cerebral lo hacen, pero no hay nada equivalente a un cubo para colocar los patrones dentro.)

Ahora nos adjudican la tarea adicional de buscar secuencias. Nos damos cuenta de que rojo-rojo-verde-púrpura-naranja-verde aparece con frecuencia. La denominamos la secuencia "rrvpnv". Hemos de advertir que sería imposible reconocer una secuencia si primero no hubiéramos clasificado los patrones en

clases diferentes. Sin haber clasificado antes cada trozo de papel en una de las diez categorías, no podríamos afirmar que dos secuencias son iguales.

Así pues, ahora estamos preparados para actuar. Vamos a mirar todos los patrones de entrada —los trozos coloreados de papel que entran desde las regiones corticales inferiores—; vamos a clasificarlos y luego a buscar secuencias. Ambos pasos, clasificación y formación de secuencias, son necesarios para crear representaciones invariables, y cada una de las regiones corticales los efectúan.

El proceso de formar secuencias da resultado cuando una entrada es ambigua, como un trozo de papel que se encuentra entre el rojo y el naranja. Tenemos que elegirle un cubo aunque no estemos seguros si es más bien rojo o más bien naranja. Si conocemos la secuencia más probable para esta serie de entradas, utilizaremos dicho conocimiento para decidir cómo clasificar la entrada ambigua. Si creemos que estamos en la secuencia "rrvpnv" porque acabamos de obtener dos rojos, un verde y un púrpura, esperaremos que el siguiente papel sea naranja. Pero llega ese siguiente papel y no es naranja, sino de un color raro entre el rojo y el naranja. Tal vez sea incluso un poco más rojo que naranja. Pero conocemos la secuencia "rrvpnv" y la esperamos, así que colocamos el papel en el cubo naranja. Utilizamos el contexto de las secuencias conocidas para resolver la ambigüedad.

Vemos que este fenómeno sucede continuamente en nuestras experiencias cotidianas. Cuando la gente habla, sus palabras particulares con frecuencia no pueden entenderse fuera de contexto. Sin embargo, cuando escuchamos una palabra ambigua dentro de una oración, esa ambigüedad no nos deja suspensos. La entendemos. De forma similar, las palabras escritas a mano a veces resultan ininteligibles fuera de contexto, pero son legibles dentro de una oración escrita completa. Las más de las veces no nos damos cuenta de que estamos completando información ambigua o incompleta de nuestras memorias de secuencias. Escuchamos lo que esperamos escuchar y vemos lo que esperamos ver, al menos cuando lo que escuchamos y vemos encaja con la experiencia pasada.

Fíjense que la memoria de las secuencias no solo nos permite resolver la ambigüedad de la entrada presente, sino también predecir qué entrada debe aparecer a continuación. Mientras su yo cortical clasifica papeles de colores, puede decir a la persona "entrada" que le pasa los papeles: "Oye, por si tienes alguna dificultad para decidir qué pasarme a continuación, según mi memoria, debe ser uno naranja". Al reconocer una secuencia de patrones, una región cortical predecirá su próximo patrón de entrada e indicará a la región inferior qué esperar.

Una región de la corteza cerebral no solo aprende secuencias conocidas, sino también cómo modificar sus clasificaciones. Digamos que comenzamos con un conjunto de cubos etiquetados "verde", "amarillo", "rojo", "púrpura" y "naranja". Estamos preparados para reconocer la secuencia "rrvpnv", así como otras

combinaciones de estos colores. ¿Pero qué ocurre si un color efectúa un cambio importante? ¿Qué sucede si cada vez que vemos la secuencia "rrvpnv" el púrpura está algo estropeado? El nuevo color se parece más al añil. Así que cambiamos el cubo púrpura para que sea el cubo "añil". Ahora los cubos encajan mejor con lo que vemos; hemos reducido la ambigüedad. La corteza cerebral es flexible.

En las regiones corticales, las clasificaciones de abajo-arriba y las secuencias de arriba-abajo interactúan de forma constante, cambiando por completo nuestras vidas. Esta es la esencia del aprendizaje. De hecho, todas las regiones de la corteza cerebral son "plásticas", lo que significa que pueden ser modificadas por la experiencia. Recordamos el mundo formando nuevas clasificaciones y secuencias.

Por último, observemos cómo estas clasificaciones y predicciones interactúan con la región superior siguiente. Otra parte de nuestra labor cortical es transmitir el nombre de la secuencia que vemos al próximo nivel superior, así que pasamos un trozo de papel con las letras "rrvpnv", que significan poco en sí mismas para esa región superior; el nombre no es más que un patrón para ser combinado con otras entradas, clasificado y luego colocado en una secuencia de orden superior. Al igual que nosotros, sigue con atención las secuencias que ve. En un momento determinado, podría decirnos: "Oye, en caso de que te cueste decidir qué pasarme a continuación, según mi memoria, predigo que debe ser la secuencia 'aarva'". Se trata, en esencia, de una instrucción sobre lo que debemos buscar en nuestro flujo de entradas. Nos esforzaremos al máximo para interpretar lo que vemos en esta secuencia.

Como mucha gente ha escuchado el término *clasificación de patrones* empleado en la inteligencia artificial y la investigación de la visión de las máquinas, observemos cómo difiere este proceso de lo que hace la corteza cerebral. Para intentar conseguir que las máquinas reconozcan objetos, los investigadores suelen crear una plantilla —digamos la imagen de una taza o alguna forma prototipo de una taza— y luego instruyen a la máquina para que casen sus entradas con la taza prototipo. Si descubre una correspondencia estrecha, el ordenador indicará que ha encontrado una taza. Pero nuestro cerebro no tiene plantillas semejantes y los patrones que recibe cada región cortical como entradas no son como fotos. No recordamos instantáneas de lo que nuestra retina ve, o instantáneas de los patrones de nuestro caracol o nuestra piel. La jerarquía de la corteza cerebral asegura que las memorias de los objetos se distribuyan por toda ella; no están localizadas en un único punto. Además, como cada región de la jerarquía forma memorias invariables, lo que una región normal de la corteza cerebral aprende son secuencias de representaciones invariables, que en sí mismas son secuencias de memorias invariables. No encontraremos una foto de una taza o de cualquier otro objeto almacenado en nuestro cerebro.

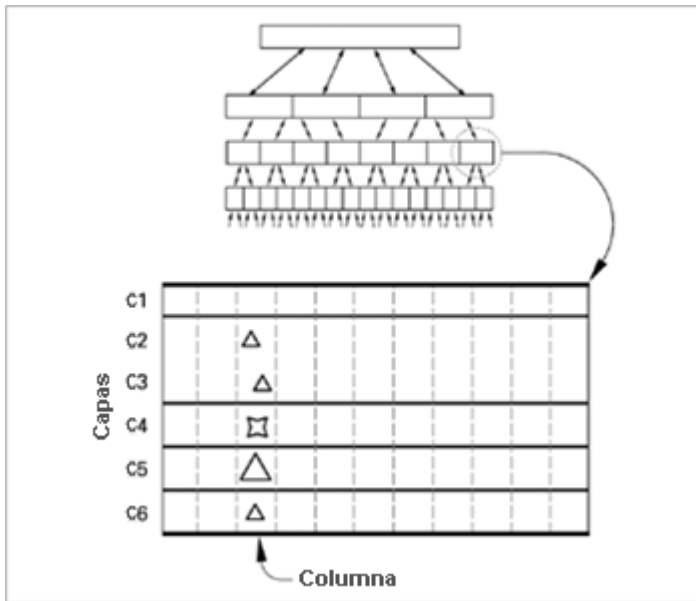
A diferencia de la memoria de una cámara, nuestro cerebro recuerda el mundo tal como es, no como aparece. Cuando pensamos sobre el mundo, recordamos

secuencias de patrones que corresponden a cómo son los objetos en el mundo y a cómo se comportan, no a cómo aparecen a través de un sentido particular en un momento determinado. Las secuencias mediante las que experimentamos los objetos del mundo reflejan la estructura invariable del mismo mundo. El orden en que experimentamos las partes del mundo está determinado por la estructura de este. Por ejemplo, podemos subirnos a un avión directamente caminando por una pasarela, pero no desde el mostrador de billetes. La secuencia por la cual experimentamos el mundo *es* la estructura real de este, y eso es lo que la corteza cerebral quiere recordar.

No obstante, no olvidemos que una representación invariable de cualquier región de la corteza cerebral puede convertirse en una predicción detallada de cómo aparecerá en nuestros sentidos transmitiendo el patrón hacia abajo de la jerarquía. De igual modo, una representación invariable en la corteza cerebral motora puede convertirse en órdenes motoras detalladas específicas de una situación, propagando el patrón hacia abajo en la jerarquía motora.

### **Aspecto de una Región de la Corteza Cerebral**

Ahora vamos a centrar nuestra atención en una región particular de la corteza cerebral, una de las cajas de la figura 5; la figura 6 muestra una región de la corteza cerebral con mayor detalle. Mi meta es enseñarles cómo las células de una región cortical pueden aprender y recordar secuencias de patrones, que es el elemento más esencial para formar representaciones invariantes y realizar predicciones. Comenzaremos con una descripción de la apariencia de una región cortical, cómo está armada. Las regiones corticales varían mucho de tamaño, siendo las mayores las áreas sensoriales primarias. V1, por ejemplo, es del tamaño aproximado de un pasaporte por el espacio que ocupa en la parte posterior del cerebro. Pero, como ya he explicado, en realidad está compuesta por muchas regiones menores del tamaño de balas de pequeño calibre. Supongamos de momento que un área cortical típica es del tamaño de una moneda pequeña.



**Figura 6. Capas y columnas en una región de la corteza cerebral.**

Pensemos en las seis tarjetas de visita que mencioné en el capítulo 3, donde cada una representa una capa diferente de tejido cortical. ¿Por qué decimos que son capas? Si tomamos nuestra región cortical del tamaño de una moneda y la colocamos bajo un microscopio, veremos que la densidad y tamaño de las células varía a medida que nos movemos de arriba abajo. Estas diferencias definen las capas. La superior, llamada capa 1, es la más característica de las seis. Tiene muy pocas células y consta en esencia de una maraña de axones que corren paralelos a la superficie cortical. Las capas 2 y 3 presentan una apariencia similar. Contienen muchas neuronas apiñadas que se denominan células piramidales porque sus cuerpos celulares se asemejan a pequeñas pirámides. La capa 4 cuenta con un tipo de células con forma de estrella. La capa 5 tiene células piramidales normales, así como una clase de células extragrandes en forma de pirámide. La capa inferior, la 6, también presenta varios tipos de neuronas únicas. Estas y otras diferencias definen a las capas.

Vemos capas horizontales, pero con mucha frecuencia los científicos hablan de columnas de células que corren perpendiculares a las capas. Se pueden concebir las columnas como “unidades” verticales de células que funcionan juntas. (El término *columna* suscita mucho debate en la comunidad de la neurociencia. Se discute su tamaño, función e importancia. Sin embargo, para nuestro objetivo podemos concebirlas en términos generales como una arquitectura columnar, pues todo el mundo está de acuerdo en que existe.) Las capas dentro de cada columna están conectadas por los axones que corren hacia arriba y hacia abajo efectuando sinapsis. Las columnas no aparecen como pequeños pilares nítidos con límites claros —nada en la corteza cerebral es tan simple—, pero cabe inferir su existencia por varias series de pruebas.



Una razón es que las células alineadas en vertical de cada columna tienden a activarse con los mismos estímulos. Si observamos con detenimiento las columnas de V1, descubrimos que algunas responden a los segmentos de línea que se inclinan en una dirección (/), y otras, a los segmentos de línea que se inclinan en otra dirección (\). Las células dentro de cada columna están muy conectadas, motivo por el cual la columna entera responde a los mismos estímulos. De forma específica, una célula activa de la capa 4 provoca que las células superiores de las capas 3 y 2 se activen, lo que luego ocasiona que las células inferiores de las capas 5 y 6 también se activen. La actividad se propaga arriba y abajo dentro de una columna de células.

Otra razón por la que hablamos de columnas tiene que ver con el modo como se forma la corteza cerebral. En un embrión, las células precursoras únicas emigran de una cavidad cerebral interior hasta el lugar donde cobra forma la corteza cerebral. Cada una de estas células se divide para crear unas cien neuronas, llamadas microcolumnas, que están conectadas en vertical del modo que acabo de describir. El término *columna* suele usarse sin demasiado rigor para describir fenómenos diferentes; puede hacer referencia a la conectividad vertical general o a grupos específicos de células del mismo progenitor. Utilizando la definición última, cabe afirmar que la corteza cerebral humana presenta una cifra aproximada de varios cientos de millones de microcolumnas.

Para lograr visualizar esta estructura columnar, imaginemos que una microcolumna tiene la anchura de un cabello humano. Tomamos miles de cabellos y los cortamos en segmentos muy reducidos —digamos de la altura de una *i* minúscula sin el punto—. Alineamos todos estos cabellos o columnas y los pegamos lado con lado como si fueran un cepillo muy denso. Luego creamos una lámina de cabellos largos y extrafinos —que representan los axones de la capa 1— y los pegamos horizontalmente en la parte superior de la lámina de cabellos cortos. Esta lámina parecida a un cepillo es un modelo simplista de nuestra región cortical del tamaño de una moneda. La información fluye en su mayor parte en la dirección de estos cabellos, horizontalmente en la capa 1 y verticalmente en las capas comprendidas entre la 2 y la 5.

Hay un detalle más de las columnas que es preciso conocer, y luego pasaremos a analizar para qué sirven. Con una inspección minuciosa, vemos que al menos el 90 por 100 de las sinapsis de las células dentro de cada columna provienen de lugares de fuera de dicha columna. Algunas conexiones llegan de columnas vecinas; otras, desde puntos equidistantes al otro lado del cerebro. Así pues, ¿cómo podemos afirmar que las columnas son importantes cuando buena parte de nuestra conexión cortical se extiende lateralmente sobre grandes áreas?

La respuesta está en el modelo de memoria-predicción. En 1979, cuando Vernon Mountcastle sostuvo que hay un solo algoritmo cortical, también propuso que la columna cortical es la unidad básica de computación de la corteza cerebral. Sin

embargo, no sabía qué funciones realiza. Yo creo que la columna es la unidad básica de predicción. Para que una columna prediga cuándo debe activarse necesita saber qué está pasando en otro lugar; de ahí las conexiones sinápticas de un lado a otro.

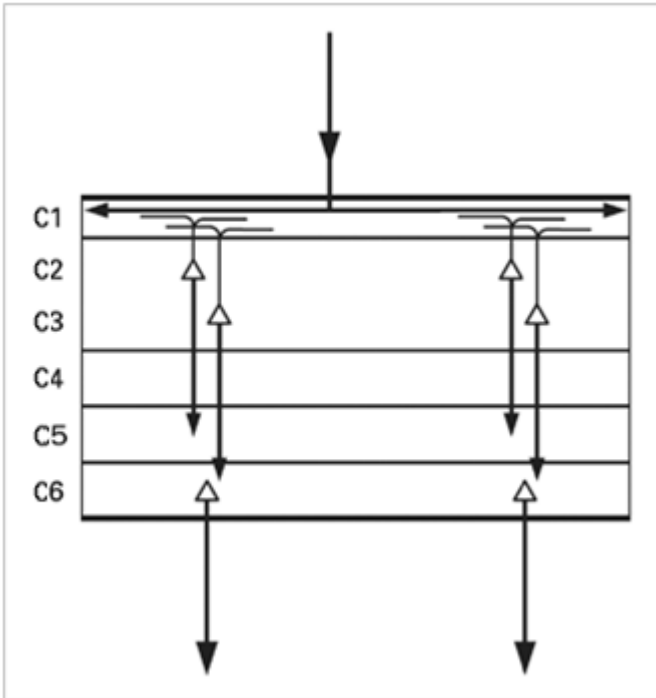
Enseguida entraremos en más detalles, pero esta es una visión preliminar para comprender por qué necesitamos ese tipo de conexión en el cerebro. Para predecir la nota siguiente de una canción necesitamos saber su nombre, en qué lugar de la canción nos encontramos, cuánto tiempo ha pasado desde la última nota y cuál fue esa última nota. El gran número de sinapsis que conectan las células de una columna con otras partes del cerebro suministra a cada una el contexto que precisa para predecir su actividad en muchas situaciones diferentes.

\* \* \*

Lo siguiente que debemos considerar es cómo estas regiones corticales del tamaño de una moneda (y sus columnas) envían y reciben información hacia arriba y abajo de la jerarquía cortical. Primero observaremos el flujo ascendente, que toma una ruta directa, representado en la figura 7. Imaginemos que contemplamos una región cortical con sus miles de columnas. Nos concentramos en una sola. Las entradas convergentes de las regiones inferiores llegan siempre a la capa 4, la principal capa de entradas. Al pasar, también forman una conexión en la capa 6 (veremos más adelante por qué es importante). Luego, las células de la capa 4 envían proyecciones hacia arriba a las células de las capas 2 y 3 de su columna. Cuando una columna proyecta información hacia arriba, muchas células de las capas 2 y 3 envían axones a la capa de entrada de la región superior siguiente. De este modo, la información fluye de región en región, ascendiendo en la jerarquía.

La información que fluye hacia abajo de la jerarquía cortical toma un camino menos directo, como se representa en la figura 8. Las de la capa 6 son las células de salida que se proyectan hacia abajo desde una columna cortical y a la capa 1 en las regiones que se encuentran por debajo en la jerarquía. En la capa 1, los axones se propagan por largas distancias en la región cortical inferior. De este modo, la información que fluye hacia abajo en la jerarquía de una columna posee el potencial de activar muchas columnas en las regiones que se encuentran por debajo de ella. En la capa 1 hay muy pocas células, pero las de las capas 2, 3 y 5 tienen dendritas en la capa 1, así que estas células pueden estimularse por las realimentaciones que recorren la capa 1. Los axones provenientes de las células de las capas 2 y 3 forman sinapsis en la capa 5 cuando abandonan la corteza cerebral, y se cree que estimulan células de las capas 5 y 6. Así pues, cabe afirmar que cuando la información fluye hacia abajo de la jerarquía presenta una ruta menos directa. Puede ramificarse en muchas direcciones diferentes mediante su propagación sobre la capa 1. La información de realimentación comienza en una célula de la capa 6 en la región superior y se extiende por la capa 1 en la región inferior. Algunas células de las capas 2, 3 y 5 de la región inferior se estimulan, y





**Figura 8. Flujo descendente de información por una región de la corteza cerebral.**

Un dato final sobre anatomía: cuando los axones abandonan la capa 6 para viajar a otros destinos, se encierran en una sustancia grasienta blanca llamada mielina. Esta denominada materia blanca se asemeja al aislamiento de un cable eléctrico de nuestra casa. Ayuda a impedir que las señales se mezclen y las hace viajar más deprisa, a velocidades que superan los 320 kilómetros por hora. Cuando los axones abandonan la materia blanca entran en una nueva columna cortical de la capa 6.

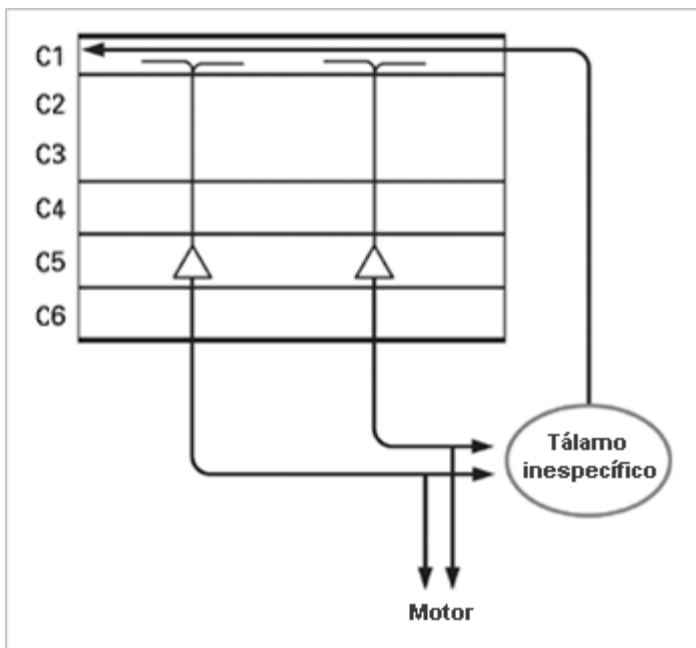
\* \* \*

**P**ara finalizar, consideremos otro método indirecto con el que las regiones corticales se comunican entre sí.

Antes de describir los detalles, quiero que recuerden las memorias autoasociativas analizadas en el capítulo 2. Como vimos, las memorias autoasociativas pueden emplearse para almacenar secuencias de patrones. Cuando la salida de un grupo de neuronas artificiales se realimenta para formar la entrada de todas las neuronas y se añade una demora a la realimentación, los patrones aprenden a seguirse en secuencia. Creo que la corteza cerebral utiliza el mismo mecanismo básico para almacenar secuencias, aunque con unos cuantos giros adicionales. En lugar de formar una memoria autoasociativa con neuronas artificiales, forma una memoria autoasociativa con columnas corticales. La salida de todas las columnas se

realimenta a la capa 1. De este modo, la capa 1 contiene información sobre qué columnas estaban activas en la región de la corteza cerebral.

Repasemos los elementos mostrados en la figura 9. Durante muchos años se ha sabido que las células particularmente grandes de la capa 5 dentro de la corteza cerebral motora (región M1) establecen contacto directo con nuestros músculos y las regiones motoras de la médula espinal. Estas células dirigen de forma literal nuestros músculos y nos hacen movernos. Siempre que hablamos, tecleamos o realizamos una conducta compleja, estas células se activan y desactivan de un modo muy coordinado que hace que nuestros músculos se contraigan.



**Figura 9. Papel de las grandes células de la capa 5 en la conducta motora.**

En fecha reciente los investigadores han descubierto que puede que las grandes células de la capa 5 desempeñen un papel en la conducta en otras partes de la corteza cerebral, no solo en las regiones motoras. Por ejemplo, las grandes células de la capa 5 de la corteza cerebral visual se proyectan a la parte del cerebro que mueve los ojos. Así pues, las áreas visuales sensoriales de la corteza cerebral, como V2 y V4, no solo procesan entradas visuales, sino que ayudan a determinar el mismo movimiento ocular y, de este modo, lo que vemos. Las grandes células de la capa 5 se ven a lo largo de toda la corteza cerebral en cada una de las regiones, lo que sugiere que tienen un papel más extenso en todo tipo de movimientos.

Además de gozar de un papel en la conducta, los axones de estas células grandes de la capa 5 pueden dividirse en dos. Una rama va a la parte del cerebro llamada

tálamo, mostrada como objeto redondo en la figura 9. El tálamo humano es del tamaño y la forma de dos pequeños huevos de pájaro. Se halla en el mismo centro del cerebro, en la cima del cerebro viejo, y está rodeado por materia blanca y la corteza cerebral. El tálamo recibe muchos axones de todas las partes de la corteza cerebral y los reenvía a esas mismas áreas. Muchos de los detalles de esas conexiones son conocidos, pero el mismo tálamo es una estructura complicada y su papel no está del todo claro. Sin embargo, es esencial para la vida normal; un tálamo dañado conduce a un estado vegetativo persistente.

Hay un par de rutas del tálamo a la corteza cerebral, pero solo una nos interesa ahora. Esta ruta comienza con las grandes células de la capa 5 que se proyectan hasta una clase de células talámicas consideradas inespecíficas. Las células inespecíficas re proyectan axones a la capa 1 sobre muchas diferentes regiones de la corteza cerebral. Por ejemplo, las células de la capa 5, a través de las regiones V2 y V4, envían axones al tálamo y este devuelve la información a las células de la capa 1 a través de V2 y V4. Otras partes de la corteza cerebral hacen lo mismo; las células de la capa 5 se proyectan a través de múltiples regiones corticales hasta el tálamo, que reenvía la información a la capa 1 de esas mismas regiones y las asociadas. Propongo que este circuito es equivalente a la realimentación diferida que permite a los modelos de la memoria asociativa aprender secuencias.

He mencionado dos entradas a la capa 1. Las regiones superiores de la corteza cerebral propagan la actividad por la capa 1 en las regiones inferiores. Las columnas activas de una región también extienden la actividad por la capa 1 en la misma región a través del tálamo. Podemos figurarnos estas entradas a la capa 1 como el nombre de una canción (entrada desde arriba) y dónde nos encontramos en una canción (actividad retardada de las columnas activas de la misma región). De este modo, la capa 1 transporta mucha de la información que necesitamos para predecir cuándo debe activarse una columna —el nombre de la secuencia y dónde nos hallamos en ella—. Utilizando estas dos señales de la capa 1, una región de la corteza cerebral es capaz de aprender y recordar múltiples secuencias de patrones.

### **Cómo Funciona una Región de la Corteza Cerebral: Los Detalles**

Con estos tres circuitos en mente —patrones convergentes ascendiendo por la jerarquía cortical, patrones divergentes descendiendo por la jerarquía cortical, y una realimentación demorada a través del tálamo— podemos comenzar a contemplar cómo una región de la corteza cerebral realiza las funciones que necesita. Lo que queremos saber es lo siguiente:

#### **1.**

¿Cómo clasifica sus entradas (como los cubos) una región de la corteza cerebral?

**2.**

¿Cómo aprende secuencias de patrones (como los intervalos de una melodía o el “ojo nariz ojo” de una cara)?

**3.**

¿Cómo forma un patrón constante o “nombre” para una secuencia?

**4.**

¿Cómo realiza predicciones específicas (aguardar el tren a la hora precisa o predecir la nota exacta de una melodía)?

Comencemos asumiendo que las columnas de una región de la corteza cerebral se parecen a los cubos que hemos empleado para clasificar nuestras entradas de papeles de colores. Cada columna representa la etiqueta de un cubo. Las células de la capa 4 de cada columna reciben fibras de entrada de varias regiones de debajo y se activarán si cuentan con la combinación adecuada de entradas. Cuando se activa una célula de la capa 4, “vota” que la entrada coincide con su etiqueta. Al igual que en la analogía de la clasificación de papeles, las entradas pueden ser ambiguas, de modo que varias columnas podrían ser correspondencias posibles para la entrada. Queremos que nuestra región cortical decida sobre una interpretación; el papel es rojo o naranja, pero no ambas cosas. Una columna con una entrada fuerte debe impedir a las demás columnas activarse.

Los cerebros cuentan con células inhibitoras que hacen justamente eso. Inhiben con fuerza a las demás neuronas vecinas de la corteza cerebral, permitiendo en la práctica que haya una ganadora. Estas células inhibitoras solo afectan al área que rodea una columna. Así pues, aunque haya una gran inhibición, muchas columnas de una región pueden seguir activas a la vez. (En los cerebros reales, nada se representa nunca con una sola neurona o una sola columna.) Para facilitar su comprensión, imaginemos que una región elige una columna ganadora. Pero recordemos en el fondo de nuestra mente que es probable que se activen muchas columnas al mismo tiempo. El proceso real que emplea una región cortical para clasificar entradas y cómo aprende a hacerlo es complicado y no se entiende bien. Para no eternizarnos con estos temas, daremos por sentado que nuestra región de la corteza cerebral ha clasificado sus entradas como actividad en un conjunto de columnas. Podemos centrarnos entonces en la formación de secuencias y nombres para estas.

¿Cómo almacena nuestra región cortical la secuencia de estos patrones clasificados? Ya he sugerido una respuesta a esta pregunta, pero ahora ahondaré en más detalles. Imaginemos que somos una columna de células y la entrada de una región inferior hace que una de las células de nuestra capa 4 se active.

Estamos contentos, y nuestra célula de la capa 4 provoca que también se activen células de las capas 2 y 3, luego de la 5 y después de la 6. La columna entera se activa cuando se penetra desde abajo. Nuestras células de las capas 2, 3 y 5 tienen cada una miles de sinapsis en la capa 1. Si algunas de estas sinapsis están activas cuando nuestras células de las capas 2, 3 y 5 se activan, las sinapsis se fortalecen. Si esto ocurre con la frecuencia necesaria, las sinapsis de la capa 1 cobran la fuerza suficiente para hacer que las células de las capas 2,3 y 5 se activen incluso cuando no se haya activado una célula de la capa 4, lo que quiere decir que partes de la columna pueden activarse sin recibir una entrada de una región inferior de la corteza cerebral. De este modo, las células de las capas 2, 3 y 5 aprenden a prever cuándo deben activarse basándose en el patrón de la capa 1. Una vez que aprende, la columna puede activarse parcialmente mediante la memoria. Cuando una columna se activa a través de las sinapsis de la capa 1, está anticipándose a ser estimulada desde abajo. Es una predicción. Si la columna pudiera hablar, diría: "Cuando he estado activa en el pasado, este conjunto particular de sinapsis de mi capa 1 ha estado activo, así que cuando vuelva a ver este conjunto particular, me activaré por adelantado".

Recordemos que la mitad de la entrada a la capa 1 proviene de las células de la capa 5 de las columnas y regiones vecinas de la corteza cerebral. Esta información representa lo que estaba sucediendo momentos antes. Representa columnas que estaban activas antes de que nuestra columna se activara. Es el intervalo previo en la melodía, o la última cosa que he visto, o la última cosa que he sentido, o el fonema anterior en el enunciado que estoy escuchando. Si el orden en el que aparecen estos patrones a lo largo del tiempo es constante, las columnas lo aprenderán. Se activarán una tras otra en la secuencia apropiada.

La otra mitad de la entrada a la capa 1 proviene de las células de la capa 6 de regiones superiores en la jerarquía. Esta información es más estacionaria. Representa el nombre de la secuencia que estamos experimentando en el momento. Si nuestras columnas son intervalos musicales, es el nombre de la melodía. Si nuestras columnas son fonemas, es la palabra hablada que estamos escuchando. Si nuestras columnas son palabras habladas, la señal de arriba es el discurso que estamos pronunciando. Así pues, la información de la capa 1 representa tanto el nombre de una secuencia como el último elemento de esta. De este modo, una columna particular puede compartirse entre muchas secuencias diferentes sin crear confusión. Las columnas aprenden a activarse en el contexto y orden precisos.

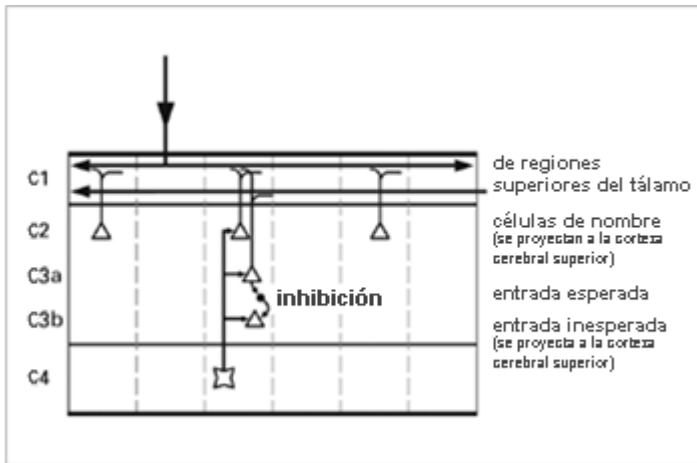
Antes de seguir avanzando, es preciso señalar que las sinapsis de la capa 1 no son las únicas que participan en el aprendizaje cuando una columna debe activarse. Como ya he mencionado, las células reciben y envían entradas a muchas columnas circundantes. Recordemos que más del 90 por 100 de las sinapsis provienen de células ajenas a la columna, y la mayoría de estas sinapsis no están en la capa 1.



Por ejemplo, las células de las capas 2, 3 y 5 tienen miles de sinapsis en la capa 1, pero también miles en sus propias capas. La idea general es que las células quieren toda información que les ayude a predecir cuándo serán estimuladas desde abajo. Por regla general, la actividad en las columnas cercanas presenta una fuerte correlación y, por lo tanto, vemos muchas conexiones directas con columnas vecinas. Por ejemplo, si una línea se mueve por nuestro campo de visión, activará columnas sucesivas. Sin embargo, la información necesaria para predecir la actividad de una columna es a menudo más global, y ahí es donde desempeñan un papel las sinapsis de la capa 1. Si fuéramos una célula o una columna, no sabríamos qué significan esas sinapsis, sino solo que nos ayudan a predecir cuándo debemos activarnos.

\* \* \*

**P**asemos ahora a analizar cómo una región de la corteza cerebral forma un nombre para una secuencia aprendida. De nuevo, imaginemos que somos una región de la corteza cerebral. Nuestras columnas activas cambian con cada nueva entrada. Hemos logrado aprender el orden en el que nuestras columnas se activan, lo que significa que algunas de las células de nuestras columnas se activan antes de la llegada de entradas provenientes de las regiones inferiores. ¿Qué información enviamos a las regiones superiores en jerarquía de la corteza cerebral? Ya hemos visto que nuestras células de las capas 2 y 3 envían sus axones a las regiones superiores siguientes. La actividad de estas células es la entrada de las regiones superiores. Pero eso es un problema. Para que la jerarquía funcione tenemos que transmitir un patrón constante durante secuencias aprendidas; hemos de pasar el nombre de una secuencia, no los detalles. Antes de aprender una secuencia podemos pasar los detalles, pero después de haberla aprendido y ser capaces de predecir qué columnas se activarán solo debemos transmitir un patrón constante. Sin embargo, cuando está en curso, todavía no hemos creado un nombre para dicha secuencia. Pasaremos todo patrón cambiante sin tener en cuenta si podemos predecirlo o no. Cuando todas las columnas se activan, sus células de las capas 2 y 3 enviarán una nueva señal que ascenderá por la jerarquía. La corteza cerebral precisa un modo de mantener constante la entrada a la región siguiente durante las secuencias aprendidas. Necesitamos un modo de desactivar la entrada de las células de las capas 2 y 3 cuando una columna predice su actividad o, de forma alternativa, mantener las células activas cuando la columna no es capaz de predecir. Este es el único modo de crear un patrón constante.



**Figura 10. Papel de las células inhibitoras en la predicción.**

No se sabe lo suficiente sobre la corteza cerebral para afirmar con exactitud cómo lo hace. Puedo imaginar varios métodos. Describiré mi favorito actual, pero no olviden que el concepto es más importante que el método específico. La creación de un patrón de "nombre" constante es un requisito de esta teoría. Todo lo que puedo mostrar en este momento es que existen mecanismos plausibles para el proceso de creación de nombres.

Imaginemos de nuevo que somos una columna, como la que se muestra en la figura 10. Queremos comprender cómo aprendemos a presentar un patrón constante a la región superior siguiente cuando logramos predecir nuestra actividad, y un patrón cambiante cuando no podemos predecirla. Comencemos asumiendo que dentro de las capas 2 y 3 hay varias clases de células. (Además de varios tipos de células inhibitoras, muchos anatomistas establecen la distinción entre tipos de células en las que denominan capas 3a y 3b, así que esta suposición es razonable.)

Asumamos también que una clase de células, llamada células de la capa 2, aprenden a permanecer activas durante secuencias aprendidas. Estas células, como grupo, representan el nombre de la secuencia. Presentarán un patrón constante a las regiones corticales superiores mientras nuestras regiones puedan predecir qué columnas se activarán a continuación. Si nuestra región de la corteza cerebral tuviera una secuencia de cinco patrones diferentes, las células de la capa 2 de todas las columnas que representan esos cinco patrones se mantendrían activas mientras estemos dentro de esa secuencia. Son el nombre de la secuencia.

Supongamos a continuación que hay otra clase de células, las de la capa 3b, que no se activan cuando nuestra columna logra predecir su entrada, pero sí lo hacen cuando no predice su actividad. Una célula de la capa 3b representa un patrón inesperado. Se estimula cuando una columna se activa de forma repentina. Se activará cada vez que una columna se estimule antes de un aprendizaje. Pero

cuando una columna aprende a predecir su actividad, la célula de la capa 3b permanece quieta. Juntas, las células de la capa 2 y 3b cumplen nuestro requisito. Antes de aprender, ambas células se activan y desactivan con la columna, pero después del entrenamiento la célula de la capa 2 está siempre activa, y la de la capa 3b, quieta.

¿Cómo aprenden a hacerlo estas células? Consideremos primero cómo desactivar a la célula de la capa 3 b cuando su columna logra predecir su actividad. Digamos que hay otra célula colocada justo encima de la célula de la capa 3b en la capa 3 a. Esta célula también tiene dendritas en la capa 1. Su única labor consiste en impedir que la célula de la capa 3 b se active cuando ve el patrón apropiado en la capa 1. Cuando la célula de la capa 3a ve el patrón aprendido en la capa 1, activa de inmediato una célula inhibidora que impide estimularse a la célula de la capa 3b. Es todo lo que se necesitaría para detener la activación de la célula de la capa 3b cuando la columna consigue predecir su actividad.

Analícemos ahora la tarea más difícil de mantener activa la célula de la capa 2 durante una secuencia conocida de patrones. Es más difícil porque un conjunto diverso de células de la capa 2 en muchas columnas diferentes necesitaría permanecer activo junto, aun cuando sus columnas particulares no estén activas. Así es como creo que podría ocurrir: las células de la capa 2 podrían aprender a ser estimuladas solo desde las regiones superiores en jerarquía de la corteza cerebral. Podrían formar sinapsis preferentes con los axones procedentes de las células de la capa 6 de las regiones más altas. De este modo, las de la capa 2 representarían el patrón de nombre constante de las regiones superiores. Cuando una región más alta de la corteza cerebral envía un patrón a la capa 1 de la región de abajo se activaría un conjunto de células de la capa 2 de la región inferior, representando a todas las columnas que son miembros de la secuencia. Puesto que estas células de la capa 2 también se re proyectan a la región superior, formarían un grupo de células semiestable. (No es probable que estas células permanezcan siempre activas; más bien se activarían de forma sincrónica, siguiendo una especie de ritmo.) Es como si la región superior enviara el nombre de una melodía a la capa 1 de debajo. Este hecho hace que se active un conjunto de células de la capa 2, una por cada una de las columnas que se activarán cuando la melodía se escuche.

La suma de todos estos mecanismos permite a la corteza cerebral aprender secuencias, efectuar predicciones y formar representaciones constantes, o "nombres" para las secuencias. Estas son las operaciones básicas para formar representaciones invariables.

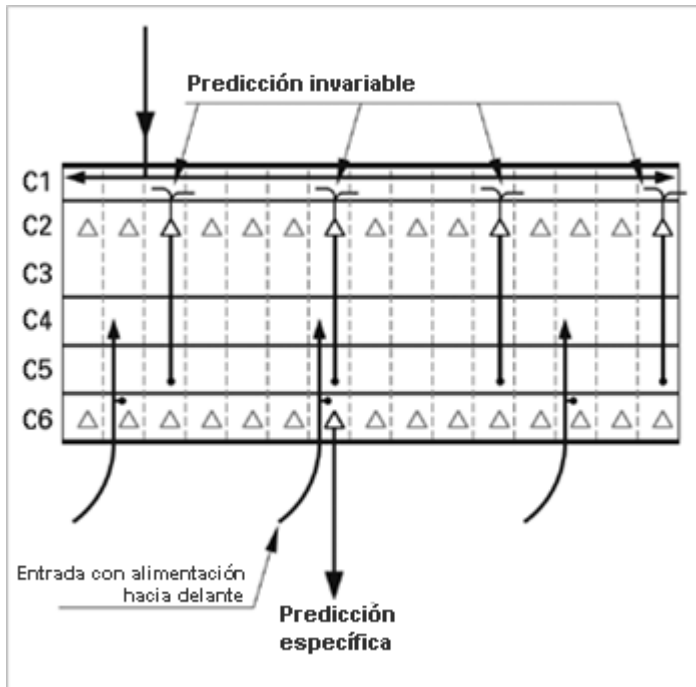
\* \* \*

¿Cómo efectuamos predicciones sobre acontecimientos que nunca hemos visto antes? ¿Cómo decidimos entre múltiples interpretaciones de una entrada? ¿Cómo

una región de la corteza cerebral realiza predicciones específicas partiendo de memorias invariables? Ya he aportado varios ejemplos, como la predicción de la nota exacta siguiente de una melodía cuando nuestra memoria no recuerda más que el intervalo entre las notas, la parábola del tren y la recitación del Discurso de Gettysburg. En estos casos, el único modo de resolver el problema es emplear la última información específica para convertir una predicción invariable en una predicción específica. Otro modo de formularlo desde el punto de vista de la corteza cerebral es afirmar que tenemos que combinar la información de alimentación hacia delante (la entrada real) con la información de realimentación (una predicción en una forma invariable).

Veamos un sencillo ejemplo de cómo creo que se realiza. Pongamos que a nuestra región de la corteza cerebral se le ha dicho que espere el intervalo musical de una quinta. Las columnas de nuestra región representan todos los intervalos específicos posibles, como do-mi, do-sol, re-la, etc. Necesitamos decidir cuál de nuestras columnas debe activarse. Cuando la región de encima nos indica que esperemos una quinta, provoca que las células de la capa 2 se activen en todas las columnas que son quintas, como do-sol, re-la y mi-si. Las células de la capa 2 de las columnas que representan otros intervalos en los que participa re, como re-mi y re-si, tienen una entrada parcial. Así pues, en la capa 2 tenemos actividad en todas las columnas que son quintas, y en la capa 4 tenemos una entrada parcial en todas las columnas que representan intervalos que incluyen re. La intersección de estos dos conjuntos representa nuestra respuesta, la columna que representa el intervalo re-la (véase figura 11).

¿Cómo encuentra la corteza cerebral esta intersección? Recordemos que ya he mencionado el hecho de que los axones de las células de las capas 2 y 3 suelen formar sinapsis en la capa 5 cuando abandonan la corteza cerebral y, de forma similar, los axones que se acercan a la capa 4 desde regiones inferiores de la corteza cerebral efectúan una sinapsis en la capa 6. La intersección de estas dos sinapsis (de arriba-abajo y de abajo-arriba) nos proporciona lo necesario. Una célula de la capa 6 que recibe estas dos entradas activas se estimulará. Una célula de la capa 6 representa lo que una región de la corteza cerebral cree que sucede, una predicción específica. Si una célula de la capa 6 pudiera hablar, tal vez diría: "Soy parte de una columna que representa el intervalo musical re-la. Otras columnas significan otras cosas. Yo hablo por mi región de la corteza cerebral. Cuando me activo, quiere decir que creo que el intervalo re-la está ocurriendo u ocurrirá. Podría activarme porque la entrada de arriba-abajo procedente de los oídos provocó que la célula de la capa 4 de mi columna excitara a la columna entera. O mi actividad tal vez signifique que hemos reconocido una melodía y estamos prediciendo este intervalo específico siguiente. De cualquier modo, mi trabajo es indicar a las regiones inferiores de la corteza cerebral qué pensamos que sucede. Yo represento nuestra interpretación del mundo, prescindiendo de si es cierta o solo imaginada".



**Figura 11. Cómo una región de la corteza cerebral efectúa predicciones específicas partiendo de memorias invariables.**

Voy a describirlo empleando otra imagen mental. Imaginemos dos trozos de papel con multitud de agujeritos perforados. Los agujeros de un papel representan las columnas que tienen células activas de la capa 2 o 3, nuestra predicción invariable. Los agujeros del otro papel representan columnas con una entrada parcial desde abajo. Si colocamos un trozo de papel encima del otro, algunos de los agujeros formarán una fila y otros no. Los agujeros que forman filas representan las columnas que pensamos que deben activarse.

Este mecanismo no solo realiza predicciones específicas, sino que también resuelve ambigüedades de las entradas sensoriales. Con mucha frecuencia la entrada a una región de la corteza cerebral será ambigua, como hemos visto con los papeles de colores, o cuando escuchamos una palabra semiincomprensible. Este mecanismo de correspondencia de abajo-arriba/arriba-abajo nos permite decidir entre dos o más interpretaciones. Y una vez que decidimos, transmitimos nuestra interpretación a la región de debajo.

En todo momento de nuestra vida consciente, cada región de la corteza cerebral está comparando un conjunto de columnas inesperadas estimuladas desde arriba con el conjunto de columnas observadas estimuladas desde abajo. Donde los dos conjuntos se intersecan es lo que percibimos. Si tuviéramos una entrada perfecta desde abajo y predicciones perfectas, el conjunto de columnas percibidas estaría contenido siempre en el conjunto de columnas predichas. Pero a menudo carecemos de tal armonía. El método de combinar la predicción parcial con la

entrada parcial resuelve la entrada ambigua, completa las partes de información que faltan y decide entre visiones alternativas. Así es como combinamos un intervalo de tono invariable con la última nota escuchada para predecir la siguiente nota específica en una melodía. Así es como decidimos si la foto es de un jarrón o de dos caras. Así es como dividimos nuestro flujo motor para escribir o pronunciar el Discurso de Gettysburg.

Por último, además de proyectarse a las regiones corticales inferiores, las células de la capa 6 pueden reenviar su salida a las células de la capa 4 de su propia columna. Cuando lo hacen, nuestras predicciones se convierten en la entrada. Es lo que hacemos cuando soñamos despiertos o pensamos. Nos permite ver las consecuencias de nuestras propias predicciones. Lo hacemos muchas horas al día cuando planeamos el futuro, ensayamos discursos y nos preocupamos por acontecimientos futuros. Stephen Grossberg, que lleva mucho tiempo modelando la corteza cerebral, lo denomina "realimentación plegada". Yo prefiero "imaginar".

\* \* \*

**U**n último tema antes de dejar este epígrafe. Ya he señalado que las más de las veces lo que vemos, escuchamos o sentimos depende mucho de nuestras propias acciones. Lo que vemos depende de dónde efectúan la sacudida ocular nuestros ojos y cómo giramos la cabeza. Lo que sentimos depende de cómo movemos las extremidades y dedos. Lo que escuchamos depende a veces de lo que decimos y hacemos.

Así pues, para predecir lo que veremos a continuación tenemos que saber qué acciones estamos acometiendo. La conducta motora y la percepción sensorial son muy interdependientes. ¿Cómo podemos hacer predicciones si lo que sentimos a continuación es en buena medida el resultado de nuestras propias acciones? Por suerte, hay una solución sorprendente y elegante para este problema, aunque muchos de los detalles no se comprenden.

El primer descubrimiento sorprendente es que la percepción y la conducta son casi idénticas. Como ya he mencionado, la mayoría —si no todas— de las regiones de la corteza cerebral, incluso las áreas visuales, participan en la creación del movimiento. Las células de la capa 5 que se proyectan al tálamo y luego a la capa 1 parecen tener también una función motora porque se proyectan de forma simultánea a áreas motoras del cerebro viejo. De este modo, se dispone del conocimiento de "lo que acaba de suceder" —tanto sensorial como motor— en la capa 1.

El segundo hecho sorprendente, y consecuencia del primero, es que la conducta motora también debe plasmarse en una jerarquía de representaciones invariables. Generamos los movimientos necesarios para llevar a cabo una acción particular, pensando en hacerla de una forma invariable al detalle. A medida que la orden

motora se desplaza jerarquía abajo, se traduce en las secuencias complejas y detalladas que se requieren para realizar la actividad que esperamos hacer. Esto sucede tanto en la corteza cerebral "motora" como "sensorial", lo que vuelve borrosa la distinción entre ambas. Si la región IT de la corteza cerebral visual percibe "nariz", el mero acto de cambiar a la representación de "ojo" generará la sacudida ocular necesaria para hacer realidad esta predicción. La sacudida ocular particular necesaria para pasar de ver una nariz a ver un ojo varía según dónde esté el rostro. Un rostro cercano requiere una sacudida ocular mayor; un rostro más distante, una sacudida ocular menor. Un rostro inclinado requiere una sacudida ocular en un ángulo diferente de otra para un rostro horizontal. Los detalles de la sacudida ocular precisa se determinan mientras la predicción de ver el "ojo" avanza hacia V1. La sacudida ocular se vuelve cada vez más específica cuanto más desciende, dando como resultado la que acaba en las fóveas justo en el blanco, o muy cerca.

Analicemos otro ejemplo. Para que yo me desplace físicamente del salón a la cocina, todo lo que mi cerebro tiene que hacer es cambiar de la representación invariable del salón a la representación invariable de la cocina. Este cambio provoca un despliegue complejo de secuencias. El proceso de generar la secuencia de predicciones de lo que veré, sentiré y escucharé mientras camino del salón a la cocina también genera la secuencia de las órdenes motoras que me hacen caminar del salón a la cocina y mover los ojos mientras lo efectúo. La predicción y la conducta motora funcionan mano a mano, mientras los patrones fluyen hacia arriba y abajo de la jerarquía cortical. Por extraño que parezca, cuando nuestra propia conducta está incluida, nuestras predicciones no solo preceden a la sensación, sino que la determinan. Pensar en pasar al patrón siguiente de una secuencia provoca una predicción en cascada de lo que debemos experimentar a continuación. A medida que se despliega la predicción en cascada, genera las órdenes motoras necesarias para cumplirla. Pensar, predecir y actuar forman parte del mismo despliegue de secuencias que descienden por la jerarquía cortical.

"Hacer" pensando, el despliegue paralelo de la percepción y la conducta motora, constituye la esencia de lo que se denomina conducta orientada a una meta. Dicha conducta es el santo grial de la robótica, y está incorporada en el tejido de la corteza cerebral.

Podemos desconectar nuestra conducta motora, por supuesto. Puedo pensar en ver algo sin verlo en realidad y puedo pensar en ir a la cocina sin hacerlo en realidad. Pero pensar en hacer algo es, de forma literal, el comienzo de nuestro modo de hacerlo.

### **Flujos Ascendentes y Descendentes**

Retrocedamos un poco para reflexionar algo más sobre cómo se mueve la información hacia arriba y abajo de la jerarquía cortical. Cuando nos desplazamos

por el mundo fluyen entradas cambiantes a las regiones inferiores de la corteza cerebral. Cada región trata de interpretar su corriente de entradas como parte de una secuencia de patrones conocida. Las columnas intentan prever su actividad. Si lo logran, pasarán un patrón estable, el nombre de la secuencia, a la región superior siguiente. De nuevo, es como si la región dijera: "Estoy escuchando una canción; este es su nombre. Puedo manejar los detalles".

¿Pero qué ocurre si llega un patrón inesperado, una nota imprevista? ¿O qué ocurre si vemos algo que no pertenece a una cara? El patrón inesperado se pasa de forma automática a la región cortical superior siguiente. Ocurre de modo natural cuando se estimulan las células de la capa 3b que no formaban parte de la secuencia esperada. La región superior puede ser capaz de comprender este nuevo patrón como la parte siguiente de su secuencia propia. Podría decir: "Veo que ha llegado una nueva nota. Tal vez es la primera de la siguiente canción del álbum. Eso parece, así que predigo que hemos pasado a la siguiente canción. Región inferior, aquí está el nombre de la siguiente canción que creo que debes estar escuchando". Pero si no ocurre este reconocimiento, el patrón inesperado continuará propagándose hacia arriba de la jerarquía cortical hasta que alguna región superior pueda interpretarlo como parte de su secuencia de hechos normal. Cuanto más necesite ascender el patrón inesperado, más regiones de la corteza cerebral participan en la resolución de la entrada inesperada. Por último, cuando una región elevada de la jerarquía piensa que puede comprender el hecho inesperado, genera una nueva predicción. Esta se propaga hacia abajo en la jerarquía hasta donde puede avanzar. Si la nueva predicción no es acertada, se detectará el error y de nuevo ascenderá por la jerarquía hasta que alguna región sea capaz de interpretarlo como parte de su secuencia activa en curso. De este modo, vemos que los patrones observados fluyen hacia arriba en la jerarquía, y las predicciones, hacia abajo. Idealmente, en un mundo conocido y predecible, la mayor parte del flujo de patrones ascendente y descendente ocurre con rapidez y en las regiones inferiores de la corteza cerebral. El cerebro trata enseguida de encontrar una parte de su modelo del mundo que sea compatible con cualquier entrada inesperada. Solo entonces puede saber razonablemente qué esperar después.

Si voy caminando por una habitación conocida de mi casa se propagarán pocos errores hacia arriba de mi corteza cerebral. Las secuencias bien aprendidas de mi casa pueden manejarse en las secciones inferiores de la jerarquía visual y motora. Conozco tan bien la habitación que incluso puedo recorrerla en la oscuridad. Mi familiaridad con el entorno libera en la práctica la mayor parte de mi corteza cerebral para otras tareas, como pensar en cerebros y escribir libros. Sin embargo, si estuviera en una habitación desconocida, sobre todo una que fuera diferente de todas las que había visto antes, no solo necesitaría mirar para ver por dónde avanzaba, sino qué patrones inesperados ascenderían cada vez más arriba de la jerarquía cortical. Cuanto menos concuerde mi experiencia sensorial con las



secuencias aprendidas, más errores surgirán. En esta situación nueva, ya no puedo pensar en cerebros porque la mayor parte de mi corteza cerebral está pendiente de los problemas de recorrer la habitación. Es una experiencia común para la gente que se baja de un avión en un país extranjero. Aunque las carreteras tal vez parezcan semejantes a las que estamos acostumbrados, puede que los coches circulen por el carril contrario, el dinero es extraño, la lengua resulta incomprensible y aprender a encontrar un baño puede precisar toda la potencia de nuestra corteza cerebral. No intente ensayar un discurso mientras camina por suelo ajeno.

La sensación de comprensión repentina, el momento de "¡ah!", puede entenderse en este modelo. Imaginemos que estamos mirando un cuadro ambiguo. Lleno de manchas de tinta y líneas dispersas, no se parece a nada. Carece de sentido. Surge la confusión cuando la corteza cerebral no puede encontrar ninguna memoria que se corresponda con la entrada. Nuestros ojos escudriñan cada lugar del cuadro. Nuevas entradas recorren todo el camino ascendente de la jerarquía cortical. Las regiones superiores comprueban muchas hipótesis diferentes, pero cuando dichas predicciones recorren el camino descendente de la jerarquía, todas y cada una resultan incompatibles con la entrada, y la corteza cerebral se ve obligada a realizar un nuevo intento. Durante este tiempo de confusión nuestro cerebro está totalmente ocupado en entender el cuadro. Por fin efectuamos una predicción de alto nivel que es la acertada. Cuando sucede esto, la predicción comienza en la cima de la jerarquía cortical y logra propagarse hasta el fondo. En menos de un segundo, a cada región se le entrega una secuencia que encaja con los datos. No ascienden más errores a la cima. Comprendemos que el cuadro se entiende; vemos un perro dálmata entre los puntos y borrones (véase figura 12).



**Figura 12. ¿Ve el dálmata?**

## ¿Puede la Realimentación Lograr Eso?

Hemos sabido durante décadas que las conexiones de la jerarquía cortical son recíprocas. Si una región A se proyecta a la región B, B se proyecta a la región A. Suele haber más fibras de axones yendo hacia delante que hacia atrás. Pero aunque esta descripción goza de una amplia aceptación, el paradigma predominante es que la realimentación desempeña un papel menor o “modulador” en el cerebro. La idea de que una señal de realimentación pueda provocar de forma inmediata y precisa que un conjunto diverso de células de la capa 2 se estimule no es la postura dominante entre los neurocientíficos.

¿Por qué debería ser así? Parte de la razón, como ya he mencionado, es que no existe una necesidad real de preocuparse por la realimentación si no se acepta el papel central de la predicción. Si se piensa que la información fluye sin interrupciones por el sistema motor, ¿por qué se necesita realimentación? Otra razón para ignorarla es que la señal de realimentación se extiende por grandes áreas de la capa 1. Por intuición se esperaría que una señal que se dispersa por una gran área tenga un efecto menor en muchas neuronas, y, de hecho, el cerebro cuenta con varias señales moduladoras semejantes que no actúan sobre neuronas específicas, sino que cambian atributos globales como el estado de alerta.

La razón final para pasar por alto la realimentación estriba en cómo creen muchos científicos que funcionan las neuronas. Las neuronas tipo cuentan con miles o cientos de miles de sinapsis. Algunas se localizan lejos del cuerpo celular; otras, justo encima, muy cerca. Las sinapsis próximas al cuerpo celular tienen una gran influencia en la estimulación de la célula. Una docena aproximada de sinapsis activas cerca del cuerpo celular pueden provocar que genere una punta o impulso de descarga eléctrica. Esto es sabido. Sin embargo, la vasta mayoría de las sinapsis no están cerca del cuerpo de la célula. Se extienden a lo largo y ancho de la estructura ramificada de las dendritas. Puesto que estas sinapsis están muy alejadas del cuerpo celular, los científicos han tendido a creer que una punta que llegue a una de estas sinapsis tendría un efecto débil o casi imperceptible en que la neurona genere una punta. El efecto de una sinapsis distante ya se habría disipado cuando llega al cuerpo celular.

Como regla general, la información que fluye hacia arriba de la jerarquía cortical se transfiere a través de las sinapsis próximas a los cuerpos celulares. De este modo, es más seguro que la información que asciende por la jerarquía pase de una región a otra. Asimismo, como regla general, la realimentación que desciende por la jerarquía cortical lo hace a través de las sinapsis alejadas del cuerpo celular. Las células de las capas 2, 3 y 5 envían dendritas a la capa 1 y forman allí muchas sinapsis. La capa 1 es una masa de sinapsis, pero todas están alejadas de los cuerpos celulares de las capas 2, 3 y 5. Además, una célula particular de la capa 2 solo formará como mucho sinapsis con una fibra de realimentación particular. Por

lo tanto, algunos científicos pueden oponerse a la idea de que un patrón breve de la capa 1 pueda provocar que un conjunto de células se estimule en las capas 2, 3 y 5. Pero esto es precisamente lo que la teoría que he expuesto requiere.

La resolución de este dilema es que las neuronas se comportan de forma diferente al modo como lo hacen en el modelo clásico. De hecho, en los años recientes, un grupo creciente de científicos ha propuesto que las sinapsis en dendritas distantes y finas pueden desempeñar un papel activo muy específico en la activación celular. En estos modelos, esas sinapsis distantes se comportan de forma diferente de las sinapsis de las dendritas más gruesas cercanas al cuerpo celular. Por ejemplo, si hubiera dos sinapsis muy cerca una de otra en una dendrita delgada, actuarían como "detectores de coincidencia". Es decir, si ambas sinapsis recibieran una punta de entrada dentro de un recuadro de tiempo reducido, podrían ejercer un gran efecto sobre la célula aun cuando estén alejadas del cuerpo celular. Podrían provocar que el cuerpo celular generara una punta. Cómo se comportan las dendritas de una neurona sigue siendo un misterio, así que no puedo decir mucho más al respecto. Lo importante es que el modelo de memoria-predicción de la corteza cerebral requiere que las sinapsis alejadas del cuerpo celular sean capaces de detectar patrones específicos.

Reflexionando a posteriori, parece casi tonto afirmar que la mayoría de las miles de sinapsis de una neurona se limitan a desempeñar un papel modulador. La realimentación masiva y el número ingente de sinapsis existen por una razón. Siguiendo esta percepción, cabe afirmar que una neurona tipo tiene la capacidad de aprender cientos de coincidencias precisas en las fibras de realimentación cuando establecen sinapsis sobre dendritas delgadas. Ello significa que cada columna de nuestra corteza cerebral es muy flexible desde la perspectiva de los patrones de realimentación que provocan que se active. Significa que cualquier rasgo particular puede asociarse de manera precisa con miles de objetos y secuencias diferentes. Mi modelo requiere que la realimentación sea rápida y precisa. Las células necesitan estimularse cuando ven un número de coincidencias precisas en sus dendritas distantes. Estos nuevos modelos de neuronas lo permiten.

### **Cómo Aprende la Corteza Cerebral**

Todas las células de todas las capas de la corteza cerebral cuentan con sinapsis, y la mayoría de estas pueden modificarse mediante la experiencia. Cabe afirmar que el aprendizaje y la memoria ocurren en todas las capas, en todas las columnas y en todas las regiones de la corteza cerebral.

Ya he hablado del libro del aprendizaje hebbiano, bautizado con ese nombre por el neuropsicólogo canadiense Donald O. Hebb. Su esencia es muy sencilla: cuando dos neuronas se activan al mismo tiempo, las sinapsis entre ellas se fortalecen. Ahora sabemos que Hebb estaba básicamente en lo cierto. Por supuesto, nada en

la Naturaleza es nunca tan sencillo, y en los cerebros reales los detalles son más complicados. Nuestros sistemas nerviosos practican muchas variaciones de la regla de aprendizaje hebbiano; por ejemplo, algunas sinapsis cambian su fuerza en respuesta a pequeñas variaciones en la sincronización de las señales neuronales; algunos cambios sinápticos son de corta duración, y otros, de larga. Pero Hebb se limitó a establecer un marco para el estudio del aprendizaje, no era una teoría final, y dicho marco ha resultado increíblemente útil.

Los principios del aprendizaje hebbiano pueden explicar la mayoría de la conducta cortical que he mencionado en este capítulo. Recordemos que ya se demostró en la década de 1970 que las memorias autoasociativas, empleando el algoritmo hebbiano clásico, son capaces de aprender patrones espaciales y secuencias de patrones. El problema principal era que las memorias no lograban manejar bien la variación. Según la teoría propuesta en este libro, la corteza cerebral ha eludido esta limitación en parte acumulando memorias asociativas en una jerarquía y en parte usando una compleja arquitectura columnar. Este capítulo se ha consagrado sobre todo a analizar la jerarquía y su funcionamiento porque es la que hace poderosa a la corteza cerebral. Así pues, en lugar de dedicarme a describir con detalles minuciosos cómo podría aprender cada célula esto o aquello, deseo abordar unos cuantos principios amplios de aprendizaje en una jerarquía.

Cuando nacemos, nuestra corteza cerebral no sabe nada. No conoce nuestra lengua, nuestra cultura, nuestra casa, nuestra ciudad, canciones, la gente con la que creceremos, nada. Toda esta información, la estructura del mundo, tiene que aprenderse. Los dos componentes básicos del aprendizaje son la clasificación de patrones y la construcción de secuencias. Estos dos componentes complementarios de la memoria interactúan. Cuando una región aprende secuencias, las entradas de las células de la capa 4 cambian. Por lo tanto, estas células de la capa 4 aprenden a formar nuevas clasificaciones, que cambian el patrón reprojectado a la capa 1, que afecta a las secuencias.

Lo básico de formar secuencias es agrupar patrones que son parte del mismo objeto. Un modo de hacerlo es agrupando los patrones que aparecen seguidos en el tiempo. Si una niña sostiene un juguete en la mano y lo mueve lentamente, su cerebro puede asumir sin duda alguna que la imagen de su retina es del mismo objeto un momento tras otro y, por lo tanto, el conjunto cambiante de patrones puede agruparse junto. En otros momentos se necesitan instrucciones externas para ayudar a decidir qué patrones deben ir juntos. Aprender que las manzanas y los plátanos son frutas, pero las nueces y el apio no, requiere un maestro que guíe a agrupar estos artículos como frutas. De cualquier modo, nuestro cerebro construye lentamente secuencias de patrones que son afines. Pero cuando una región de la corteza cerebral construye patrones, la entrada en la región siguiente cambia. La entrada pasa de representar en su mayoría patrones particulares a representar grupos de patrones. La entrada de una región cambia de notas a

melodías, de letras a palabras, de narices a rostros, y así sucesivamente. Como las entradas de abajo-arriba de una región se vuelven más “orientadas al objeto”, la región superior de la corteza cerebral puede ahora aprender secuencias de esos objetos de orden superior. Donde antes una región construía secuencias de letras, ahora construye secuencias de palabras. El resultado inesperado de este proceso es que durante el aprendizaje repetitivo las representaciones de objetos descienden por la jerarquía cortical. Durante los primeros años de nuestra vida, nuestras memorias del mundo se forman primero en las regiones superiores de la corteza cerebral, pero a medida que aprendemos se reforman en partes cada vez más bajas de la jerarquía cortical. No es que el cerebro las mueva; tiene que reaprenderlas una y otra vez. (No sugiero que todas las memorias comiencen en la cima de la corteza cerebral. La formación real de las memorias es más compleja. Creo que la clasificación de patrones de la capa 4 empieza abajo y va ascendiendo. Pero a medida que lo hace, comenzamos a formar secuencias que luego descienden. Es la *memoria de las secuencias* la que sugiero que se reforma cada vez más abajo de la corteza cerebral.) A medida que las representaciones sencillas van descendiendo, las regiones más elevadas son capaces de aprender patrones más complejos y sutiles.

Se puede comprobar la creación y movimiento descendente de la memoria jerárquica observando cómo aprende un niño. Analicemos cómo aprendimos a leer. Lo primero que aprendemos es a reconocer letras impresas individuales. Es una tarea lenta y difícil que requiere un esfuerzo consciente. Luego pasamos a reconocer palabras sencillas. De nuevo, resulta difícil y lento al principio, incluso para palabras de tres letras. El niño puede leer cada letra en sucesión y pronunciar las letras una tras otra, pero necesita mucha práctica antes de ser capaz de reconocer la palabra en sí como tal. Tras aprender palabras sencillas, luchamos con palabras multisilábicas complicadas. Al principio pronunciamos cada sílaba concatenándolas como hicimos con las letras cuando aprendimos palabras sencillas. Después de años de práctica, una persona puede leer deprisa. Llegamos a un punto en el que no vemos todas las letras individuales, sino que reconocemos palabras completas y a menudo frases completas de una ojeada. No es solo que seamos más rápidos; estamos reconociendo las palabras y frases como entidades. Cuando leemos una palabra completa de una vez, ¿seguimos viendo las letras? Sí y no. Es evidente que la retina ve las letras y, por lo tanto, también lo hacen las regiones de V1. Pero su reconocimiento ocurre bastante abajo en la jerarquía cortical. Digamos en V2 o V4. Cuando la señal llega a IT, las letras individuales ya no están representadas. Lo que al principio costaba el esfuerzo de la corteza cerebral visual completa —reconocer letras individuales—, ahora ocurre más cerca de la entrada visual. A medida que la memoria de objetos sencillos como letras desciende por la jerarquía, las regiones superiores tienen la capacidad de aprender objetos complejos como palabras y frases.

Aprender a leer música es otro ejemplo. Al principio hay que concentrarse en cada nota. Con la práctica, se empieza a reconocer secuencias de notas comunes, luego frases completas. Tras mucha práctica, es como si no se viera la mayoría de las notas. La música de la partitura está ahí solo para recordar la estructura principal de la pieza; las secuencias detalladas se han memorizado más abajo. Este tipo de aprendizaje ocurre tanto en las áreas motoras como sensoriales.

Un cerebro joven es más lento para reconocer entradas y efectuar órdenes motoras porque las memorias usadas en estas tareas se encuentran más arriba de la jerarquía cortical. La información tiene que fluir el camino completo arriba y abajo, tal vez con múltiples pases, para resolver los conflictos. Lleva tiempo que las señales neuronales recorran arriba y abajo la jerarquía cortical. Un cerebro joven tampoco ha formado todavía secuencias complejas en la cima y, por lo tanto, no puede reconocer y reenviar patrones complejos. No es capaz de entender la estructura de orden superior del mundo. Comparados con los de un adulto, el lenguaje del niño es sencillo, su música es sencilla y sus interacciones sociales son sencillas.

Si estudiamos un conjunto particular de objetos una y otra vez, nuestra corteza cerebral reforma las representaciones de la memoria de esos objetos jerarquía abajo, lo que libera a la cima para aprender relaciones más sutiles y complejas. Según la teoría, esto es lo que hace un experto.

En mi trabajo de diseño de ordenadores, algunas personas se sorprenden de la rapidez con la que puedo mirar un producto y ver los problemas inherentes en su diseño. Tras veinticinco años de diseñar ordenadores, cuento con un modelo superior a la media de los temas asociados con los aparatos informáticos portátiles. De modo similar, un padre con experiencia reconoce con facilidad por qué está disgustado su hijo, mientras que a otro primerizo tal vez le cueste manejar la situación. Un director empresarial puede ver con facilidad los fallos y ventajas de la estructura de una organización, mientras que el director primerizo no alcanza a comprender esas cosas. Tienen la misma entrada, pero el modelo del novato no es tan sofisticado. En todos estos casos y en miles más comenzamos aprendiendo lo básico, la estructura más sencilla. Con el tiempo nuestro conocimiento desciende por la jerarquía cortical y, por lo tanto, en la cima tenemos la oportunidad de aprender estructuras de orden superior. Es esta estructura de orden superior la que nos otorga experiencia. Los expertos y genios poseen cerebros que ven la estructura de la estructura y los patrones de los patrones más allá que los demás. Podemos convertirnos en expertos mediante la práctica, pero sin duda existe también un componente genético.

### **El Hipocampo: en la Cima de Todo**

Tres grandes estructuras cerebrales se encuentran bajo la lámina de la corteza cerebral y se comunican con ella. Son los núcleos basales, el cerebelo y el

hipocampo. Los tres existían antes que la corteza cerebral. En líneas muy generales, se puede decir que los núcleos basales eran el sistema motor primitivo, el cerebelo aprendía la sincronización precisa de las relaciones de acontecimientos y el hipocampo almacenaba memorias de acontecimientos y lugares específicos. Hasta cierto punto, la corteza cerebral ha subsumido sus funciones originales. Por ejemplo, un ser humano nacido sin buena parte del cerebelo tendrá deficiencias en sincronización y deberá aplicar un esfuerzo más consciente cuando se mueve, pero por lo demás será bastante normal.

Sabemos que la corteza cerebral es responsable de todas las secuencias motoras complejas y puede controlar directamente nuestras extremidades. No es que los núcleos basales carezcan de importancia, pero la corteza cerebral ha asumido una gran parte del control motor. Debido a ello, he descrito su función general independiente de los núcleos basales y el cerebelo. Puede que algunos científicos no estén de acuerdo con este planteamiento, pero es el que he usado en este libro y mi trabajo.

Sin embargo, el hipocampo es harina de otro costal. Es una de las áreas del cerebro más estudiadas porque resulta esencial para la formación de nuevas memorias. Si perdemos ambas mitades del hipocampo (al igual que muchas partes del sistema nervioso, existe en el lado izquierdo y derecho del cerebro) nos quedamos sin facultad para formar la mayoría de las nuevas memorias. Sin el hipocampo somos capaces de seguir hablando, caminando, viendo y escuchando, y durante un breve tiempo pareceremos casi normales, pero en realidad estamos profundamente deteriorados: no podemos recordar nada nuevo. Nos acordamos de un amigo que conocimos antes de perder el hipocampo, mas no logramos recordar a una persona nueva. Aunque nos reunamos con nuestro médico cinco veces al día durante un año, cada vez sería como la primera. No tendríamos memoria de los hechos que ocurrieron tras la pérdida del hipocampo.

Durante muchos años detesté reflexionar sobre el hipocampo porque carecía de sentido para mí. Sin duda, es esencial para el aprendizaje, pero no es el depositario supremo de la mayor parte de lo que conocemos. Lo es la corteza cerebral. La visión clásica del hipocampo es que allí se forman memorias nuevas que más tarde, a lo largo de un período de días, semanas o meses, se transfieren a la corteza cerebral. Esto no tenía sentido para mí. Sabemos que la vista, el sonido y el tacto —nuestra corriente de datos sensoriales— fluyen directamente a las áreas sensoriales de la corteza cerebral sin pasar primero por el hipocampo. Me parecía que esta información sensorial debía formar de modo automático nuevas memorias en la corteza cerebral. ¿Por qué necesitamos un hipocampo para aprender? ¿Cómo podía una estructura separada como el hipocampo interferir e impedir el aprendizaje en la corteza cerebral si se limitaba volver a transferir la información a la corteza cerebral?

Decidí dejar de lado el hipocampo, figurándome que algún día su papel se aclararía. Ese día llegó a finales de 2002, más o menos por la época en la que comencé a escribir este libro. Uno de mis colegas del Redwood Neuroscience Institute, Bruno Olshausen, señaló que las conexiones entre el hipocampo y la corteza cerebral sugieren que el primero es la región suprema de la corteza cerebral y no una estructura separada. Según este planteamiento, el hipocampo ocupa el pico de la pirámide cortical, el bloque superior de la figura 5. La corteza cerebral apareció en el escenario evolutivo emparedada entre el hipocampo y el resto del cerebro. Al parecer, este planteamiento del hipocampo como la cima de la jerarquía cortical se conocía desde hacía algún tiempo, pero yo no estaba al corriente. Había hablado con varios expertos y les había pedido que me explicaran cómo esta estructura con forma de silla de montar podía transferir memorias a la corteza cerebral. Ninguno supo hacerlo, ni tampoco mencionó que se encontrara en la cúspide de la pirámide cortical, debido probablemente a que no solo se encuentra en dicha posición, sino que se conecta de forma directa con muchas partes más antiguas del cerebro.

No obstante, percibí al instante que esta nueva perspectiva era la solución a mi confusión.

Reflexionemos sobre la información que fluye de los ojos, oídos y piel a la corteza cerebral. Cada región de esta trata de entender lo que dicha información significa. Cada región intenta comprender la entrada en virtud de las secuencias que conoce. Si comprende la entrada, dice: "Lo entiendo, es parte del objeto que estoy viendo. No pasaré los detalles". Si una región no comprende la entrada en curso, la pasa jerarquía arriba hasta que alguna región superior sí lo hace. Sin embargo, un patrón verdaderamente nuevo escalará cada vez más arriba de la jerarquía. Cada región más elevada dice: "No sé qué es esto, no lo he previsto; ¿por qué no lo revisáis más arriba?". El efecto neto es que cuando llegamos a la cima de la pirámide cortical lo que hemos dejado es información que no puede comprenderse con la experiencia previa. Nos queda la parte de la entrada que es verdaderamente nueva e inesperada.

En un día normal nos encontramos muchas cosas nuevas que llegan hasta la cima —por ejemplo, un relato en el periódico, el nombre de la persona que hemos conocido esta mañana y el accidente de coche que hemos visto camino a casa—. Son estos residuos inexplicados e inesperados, las cosas nuevas, lo que entra en el hipocampo y se almacena allí. Esta información no se guardará para siempre. Será retransferida a la corteza cerebral o acabará perdiéndose.

Me he dado cuenta de que, a medida que voy cumpliendo años, me cuesta recordar cosas nuevas. Por ejemplo, mis hijos recuerdan los detalles de la mayoría de las obras de teatro que han visto el año pasado. Yo no. Tal vez sea porque he visto tantas en mi vida que rara vez me encuentro algo realmente nuevo. Las nuevas obras encajan en las memorias de las obras pasadas y la información no



llega a mi hipocampo. Para mis hijos, cada obra es una novedad y sí llega al hipocampo. Si esto es cierto, cabría afirmar que cuanto más sabemos, menos recordamos.

A diferencia de la corteza cerebral, el hipocampo posee una estructura heterogénea con varias regiones especializadas. Realiza a la perfección la única tarea de almacenar enseguida cualquier patrón que ve. Se encuentra en la posición perfecta, en la cima de la pirámide cortical, para recordar lo que es nuevo. También está en la posición perfecta para recordar esas nuevas memorias, permitiendo que se almacenen en la jerarquía cortical, que es un proceso algo más lento. En el hipocampo podemos recordar al instante un acontecimiento nuevo para él, pero solo recordaremos algo de forma permanente en la corteza cerebral si lo experimentamos una y otra vez, ya sea en la realidad o pensando en ello.

### **Una Ruta Alternativa para Ascender en la Jerarquía**

Nuestra corteza cerebral cuenta con una segunda ruta importante para pasar información de una región a otra ascendiendo en la jerarquía. Esta ruta alternativa comienza con las células de la capa 5 que se proyectan en el tálamo (una parte diferente de la que hemos analizado antes) y luego desde este a las regiones superiores de la corteza cerebral. Siempre que dos regiones de la corteza cerebral se conectan directamente entre sí de forma jerárquica, también lo hacen indirectamente a través del tálamo. Esta segunda ruta solo pasa información jerarquía arriba, no abajo. Así pues, cuando ascendemos por la jerarquía cortical, existe un camino directo entre dos regiones y otro indirecto a través del tálamo.

El segundo camino presenta dos modos de operación determinados por las células del tálamo. En un modo, la ruta está en buena parte cerrada, así que la información no fluye a través de ella. En el otro modo, la información fluye con precisión entre las regiones. Dos científicos, Murray Sherman, de la Universidad Estatal de Nueva York, en Stony Brook, y Ray Guillery, de la Escuela de Medicina de la Universidad de Wisconsin, han descrito esta ruta alternativa y propugnado que podría ser tan importante como la directa (tal vez más) que ha constituido el tema de este capítulo hasta el momento. Tengo una hipótesis sobre lo que hace esta segunda ruta.

Lean esta palabra: *imaginación*. La mayoría de la gente puede reconocerla con una sola mirada, una fijación. Ahora miren la letra *i* en medio de la palabra. Ahora miren el punto sobre la *i*. Sus ojos pueden fijarse en el mismo emplazamiento exacto, pero en un caso ven la palabra; en el siguiente, la letra, y en el último, el punto. Si les cuesta, traten de decir "punto", "i" e "imaginación" mientras miran fijamente el punto. En todos los casos entra la misma información exacta en V1, pero cuando llega a una región superior como IT perciben diferentes cosas, diferentes grados de detalle. La región IT sabe cómo reconocer los tres objetos. Puede reconocer el punto aislado en la letra *i* y en la palabra completa de un

vistazo. Pero cuando perciben la palabra completa, V2, V4 y V1 se ocupan de los detalles, y todo lo que llega a conocer IT al respecto es la palabra. Por regla general, no percibimos las letras individuales mientras leemos; percibimos palabras o frases. Pero podemos hacerlo si queremos. Hacemos esta especie de cambio de atención continuamente, mas no solemos tener conciencia de ello. Puedo estar escuchando música de fondo y apenas darme cuenta de la melodía, pero si lo intento, soy capaz de aislar al cantante o al bajo. El mismo sonido entra en mi cabeza, pero puedo centrar mis percepciones. Cada vez que nos rascamos la cabeza, el movimiento provoca un fuerte sonido interno, pero no solemos percibirlo. Sin embargo, si nos centramos en él, logramos escucharlo con claridad. Este es otro ejemplo de la entrada sensorial que por lo general se maneja en regiones bajas de la jerarquía cortical, pero que puede ascender a niveles superiores si le prestamos atención.

Mi conjetura es que la ruta alternativa por el tálamo es el mecanismo mediante el cual atendemos a los detalles que normalmente no percibiríamos. Sorteamos el agrupamiento de secuencias en la capa 2 y envía los datos brutos a la siguiente región superior de la corteza cerebral. Los biólogos han mostrado que la ruta alternativa puede abrirse de dos modos. Uno es el método que emplearon cuando les pedí que prestaran atención a detalles que no suelen percibir, como el punto sobre la letra *i* o el sonido al rascarse la cabeza. El segundo método que puede activar esta ruta es una gran señal inesperada de abajo. Si la entrada en la ruta alternativa tiene la fuerza suficiente, envía una señal de alerta a la región superior, que puede abrir de nuevo la ruta. Por ejemplo, si les mostrara una cara y les preguntara qué era, dirían: "Cara". Si les mostrara la misma cara pero tuviera una extraña cicatriz en la nariz, primero la reconocerían, pero luego sus niveles inferiores de visión se darían cuenta de que algo va mal. Este error obliga a abrir la ruta de la atención. Los detalles tomarán ahora el camino alternativo, sorteando el agrupamiento que ocurre normalmente, y la marca suscitaría su atención. Ahora ven la cicatriz y no solo la cara. Si fuera lo bastante rara, la cicatriz podría ocupar su atención completa. De este modo, los acontecimientos poco habituales suscitan de inmediato su atención. Este es el motivo por el que no podemos evitar fijarnos en las deformidades y otros patrones inusitados. Nuestro cerebro lo hace de forma automática. Sin embargo, a menudo los errores no tienen la fuerza necesaria para abrir la ruta alternativa. Por ello, a veces no nos damos cuenta de si una palabra está mal escrita cuando la leemos.

## **Reflexiones Finales**

Para encontrar y establecer un nuevo marco científico es necesario buscar los conceptos más sencillos capaces de unir y explicar lo que eran grandes cantidades de datos dispares. Una consecuencia inevitable de este proceso es que el péndulo se incline demasiado hacia la simplificación. Es probable que se pasen por alto detalles importantes y que se malinterpreten datos. Si el marco se afianza, es

inevitable que surjan ajustes que muestren dónde fue demasiado lejos la propuesta inicial, dónde se quedó corta o dónde estaba el error.

En este capítulo he presentado muchas ideas especulativas sobre el funcionamiento de la corteza cerebral. Cuento plenamente con que varias de dichas ideas resulten erróneas, y es muy probable que todas sean revisadas. Hay además muchos detalles que ni siquiera he mencionado. El cerebro es muy complejo; los neurocientíficos que lean este libro sabrán que he presentado una burda caracterización de las complejidades de un cerebro real. No obstante, creo que el marco en su conjunto es sólido. No me queda más que esperar que las ideas centrales se conserven cuando los detalles cambien frente a nuevos datos y comprensión.

Por último, tal vez les cueste aceptar la idea de que un sistema de memoria sencillo pero grande pueda dar como resultado todo lo que hacen los humanos. ¿Sería posible que ustedes y yo no fuéramos más un sistema de memoria jerárquico? ¿Podrían almacenarse nuestras vidas, credos y ambiciones en miles de billones de sinapsis diminutas? En 1984 comencé a escribir programas informáticos de manera profesional. Había escrito pequeños programas antes, pero era la primera vez que programaba un ordenador con una interfaz gráfica de usuario, y la primera vez que trabajaba en aplicaciones grandes y complicadas. Escribía *software* para un sistema operativo creado por Grid Systems. Con ventanas, fuentes múltiples y menús, el sistema operativo de Grid era muy avanzado para su época.

Un día se me ocurrió que lo que hacía rayaba en lo imposible. Como programador escribía una por una líneas de código. Agrupaba las líneas de código en bloques llamados subrutinas. Las subrutinas se agrupaban en módulos. Los módulos se combinaban para formar una aplicación. Los programas de hoja de cálculo en los que estaba trabajando tenían tantas subrutinas y módulos que ninguna persona sería capaz de entenderlo todo. Era complejo. No obstante, una sola línea de código hacía muy poco. Colocar un píxel en el visualizador llevaba varias líneas de código. Sacar una pantalla completa para la hoja de cálculo requería que el ordenador ejecutara millones de instrucciones que se extendían en cientos de subrutinas. Las subrutinas exigían otras subrutinas de forma repetitiva. Era tan complicado que resultaba imposible saber todo lo que pasaría cuando el programa estuviera en funcionamiento. Se me ocurrió que era muy poco probable que cuando el programa funcionara obtuviera su imagen en lo que parecería un instante. Su apariencia externa eran tablas de números, etiquetas, texto y gráficos. Se comportaba como una hoja de cálculo. Pero yo sabía lo que pasaba dentro del ordenador, cuyo procesador ejecutaba una a una instrucciones sencillas. Era difícil creer que pudiera abrirse paso por el laberinto de módulos y subrutinas, y ejecutar todas esas instrucciones tan deprisa. Si no lo hubiera conocido tan bien, habría estado seguro de que el conjunto no podría funcionar. Me daba cuenta de que si

alguien hubiera inventado el concepto de ordenador con interfaz gráfica de usuario y una aplicación de hoja de cálculo, y me lo hubiera presentado sobre el papel, lo habría rechazado como algo irreal. Habría afirmado que tardaría una eternidad en hacer cualquier cosa. Era un pensamiento humillante, porque sí funcionaba. Fue entonces cuando me di cuenta de que mi sentido intuitivo sobre la velocidad del microprocesador y la potencia del diseño jerárquico era inadecuado.

Hay en ello una lección sobre la corteza cerebral. No está formada por componentes superrápidos y las reglas con las que opera no son tan complejas. Sin embargo, sí presenta una estructura jerárquica que contiene miles de millones de neuronas y billones de sinapsis. Si nos cuesta imaginar cómo ese sistema de memoria, sencillo desde la perspectiva lógica pero vasto desde la numérica, puede crear nuestra conciencia, lenguajes, culturas, arte, este libro y nuestra ciencia y tecnología, sugiero que se debe a que nuestro sentido intuitivo sobre la capacidad de la corteza cerebral y la potencia de su estructura jerárquica es inadecuado. Y, como un ordenador, acabaremos construyendo máquinas inteligentes que funcionen con los mismos principios.

## Conciencia y Creatividad

**Cuando** doy charlas sobre mi teoría del cerebro, el público suele captar enseguida el significado de la predicción como algo ligado con multitud de actividades humanas, y me formulan preguntas relacionadas. ¿De dónde proviene la creatividad? ¿Qué es la conciencia? ¿Qué es la imaginación? ¿Cómo se puede separar la realidad de las falsas creencias? Aunque estos temas no han estado entre mis motivaciones prioritarias para estudiar los cerebros, son de interés para casi todo el mundo. No pretendo ser un experto en ellos, pero el marco de memoria-predicción de la inteligencia puede aportar algunas respuestas y percepciones útiles. En este capítulo abordo algunas de las preguntas más frecuentes.

### ¿Son Inteligentes los Animales?

¿Es una rata inteligente? ¿Es un gato inteligente? ¿Cuándo comenzó la inteligencia en el período evolutivo? Me gustan estas preguntas porque la respuesta me resulta sorprendente.

Todo lo que he escrito hasta el momento sobre la corteza cerebral y su funcionamiento se basa en una premisa muy básica: el mundo posee una estructura y, por lo tanto, es predecible. Hay patrones en el mundo: las caras tienen ojos, los ojos tienen pupilas, los fuegos son calientes, la gravedad hace que los objetos caigan, las puertas se abren y se cierran, y así sucesivamente. El mundo no es aleatorio, ni tampoco homogéneo. La memoria, la predicción y la conducta no tendrían sentido si el mundo careciera de estructura. Toda conducta, sea la de un ser humano, un caracol, un organismo unicelular o un árbol, es un medio de explotar la estructura del mundo en beneficio de la reproducción.

Imaginemos un organismo unicelular que vive en una charca. La célula tiene un flagelo que le permite nadar. En la superficie de la célula hay moléculas que detectan la presencia de nutrientes. Como no todas las zonas de la charca presentan la misma concentración de nutrientes, se da un cambio gradual en el valor —o gradiente— de los nutrientes de un lado de la célula al otro. Cuando nada por la charca, la célula puede detectar el cambio. Es una forma de estructura sencilla en el mundo del animal unicelular. La célula explota su percepción química nadando hacia los lugares con concentraciones más altas e nutrientes. Cabría decir que este organismo simple está efectuando una predicción. Predice que nadando de cierto modo hallará más nutrientes. ¿Toma parte la memoria en esta predicción? Así es. La memoria está en el ADN del organismo. El animal unicelular no aprendió en su vida a explotar este gradiente, sino que el aprendizaje ocurrió en el período evolutivo y está almacenado en su ADN. Si la estructura del mundo cambiara de repente, este animal unicelular particular no aprendería a adaptarse.

No podría alterar su ADN o la conducta resultante. Para esta especie el aprendizaje solo ocurre mediante procesos evolutivos a lo largo de muchas generaciones.

¿Es inteligente este organismo unicelular? Utilizando la noción cotidiana de inteligencia humana, la respuesta es no. Pero el animal sí se encuentra en el extremo lejano de un continuo de especies que usan la memoria y la predicción para lograr reproducirse mejor, y según este rasero más académico la respuesta es sí. No se trata de etiquetar a algunas especies como inteligentes y a otras como no inteligentes. Todos los seres vivos usan la memoria y la predicción. No hay más que un continuo de métodos y complejidad en su modo de hacerlo.

Las plantas también emplean la memoria y la predicción para explotar la estructura del mundo. Un árbol efectúa una predicción cuando envía sus raíces bajo el suelo y sus ramas y hojas hacia el cielo. El árbol predice dónde va a encontrar agua y minerales basándose en la experiencia de sus antepasados. Por supuesto, un árbol no piensa; su conducta es automática. Pero la especie explota la estructura del mundo del mismo modo que el organismo unicelular. Cada especie de plantas posee un conjunto característico de conductas que explotan partes ligeramente diferentes de la estructura del mundo.

Las plantas acabaron desarrollando sistemas de comunicación basados en buena medida en la liberación lenta de señales químicas. Si un insecto daña una parte de un árbol, este envía productos químicos a través de su sistema vascular a otras partes, lo que desencadena un sistema de defensa como la fabricación de toxinas. Mediante ese sistema de comunicación, el árbol puede mostrar una conducta algo más compleja. Es probable que las neuronas evolucionaran como un medio de comunicar información más deprisa que un sistema vascular vegetal. Cabría concebir una neurona como una célula con sus propios apéndices vasculares. En un momento determinado, en lugar de trasladar lentamente productos químicos a través de esos apéndices, la neurona comenzó a utilizar puntas electroquímicas que viajan con mucha mayor velocidad. Es probable que al principio la transmisión sináptica rápida y los sistemas nerviosos sencillos no supusieran mucho aprendizaje. El juego consistía solo en señalar con mayor rapidez.

Pero luego, en el transcurso del período evolutivo, sucedió algo muy interesante. Las conexiones entre las neuronas se hicieron modificables. Una neurona podía enviar o no una señal dependiendo de lo que hubiera pasado hacía poco. Ahora la conducta podía modificarse en el curso de la vida de un organismo. El sistema nervioso se hizo plástico, al igual que la conducta. Como se podían formar memorias con rapidez, el animal era capaz de aprender la estructura de su mundo durante su propia vida. Si el mundo cambiara de improviso —digamos que un nuevo depredador entrara en escena—, el animal no tendría que proseguir con su conducta determinada genéticamente, pues tal vez ya no fuera la apropiada. Los sistemas nerviosos plásticos se convirtieron en una ingente ventaja evolutiva que

llevó al surgimiento de nuevas especies, de los peces a los caracoles y los humanos.

Como hemos visto en el capítulo 3, todos los mamíferos poseen un cerebro viejo, encima del cual se asienta la corteza cerebral, que no es más que el tejido neuronal de evolución más reciente. Pero con su estructura jerárquica, representaciones invariables y predicciones por analogía, la corteza cerebral permite a los mamíferos explotar mucho más la estructura del mundo de lo que es capaz un animal que no la posea. Nuestros antepasados dotados de corteza cerebral podían imaginar cómo hacer una red para atrapar peces. Los peces no son capaces de aprender que las redes significan muerte ni de idear cómo construir herramientas para cortarlas. Todos los mamíferos, de las ratas a los gatos y los humanos, tienen corteza cerebral. Todos son inteligentes, pero en grados diferentes.

### **¿Qué es Diferente en la Inteligencia Humana?**

El marco de memoria-predicción ofrece dos respuestas a esta pregunta. La primera es bastante directa: nuestra corteza cerebral es mayor que la del mono o el perro, por ejemplo. Al ampliar la lámina cortical hasta el tamaño de una gran servilleta, nuestros cerebros pueden aprender un modelo más complejo del mundo y efectuar predicciones más complicadas. Vemos analogías más profundas, más estructura en la estructura, que otros mamíferos. Si deseamos encontrar un compañero, no nos limitamos a mirar atributos sencillos como la salud, sino que entrevistamos a sus amigos y padres, observamos cómo se conducen y hablan, y juzgamos su honradez. Observamos estos atributos secundarios y terciarios para tratar de predecir cuál será la conducta de nuestro compañero en el futuro. Los corredores de Bolsa buscan una estructura en los patrones del mercado. Los matemáticos buscan una estructura en los números y ecuaciones. Los astrónomos buscan una estructura en los movimientos de los planetas y las estrellas. Nuestra corteza cerebral más grande nos permite contemplar nuestra casa como parte de una ciudad, que es parte de una región, que es parte de un planeta, que es parte de un gran Universo: la estructura dentro de la estructura. Ningún otro mamífero puede rumiar a esta profundidad. Estoy seguro de que mi gata no tiene un concepto del mundo fuera de nuestra casa.

La segunda diferencia entre los seres humanos y los restantes mamíferos es que tenemos lenguaje. Se han escrito libros enteros sobre las supuestas propiedades únicas del lenguaje y cómo se desarrolló. Sin embargo, encaja a la perfección en el marco de la memoria-predicción sin ninguna salsa especial ni maquinaria específica. Las palabras habladas y escritas no son más que patrones del mundo, al igual que las melodías, los coches y las casas. La sintaxis y la semántica del lenguaje no son diferentes de la estructura jerárquica de otros objetos cotidianos. Y del mismo modo que asociamos el sonido de un tren con la imagen de la memoria visual de un tren, asociamos las palabras habladas con nuestra memoria

de sus homólogos físicos y semánticos. Mediante el lenguaje un ser humano puede evocar recuerdos y crear nuevas yuxtaposiciones de objetos mentales en otro ser humano. El lenguaje es analogía pura, y por su mediación podemos lograr que otros seres humanos experimenten y aprendan cosas que tal vez no hayan visto nunca. El desarrollo del lenguaje requirió una gran corteza cerebral capaz de manejar la estructura nido de la sintaxis y la semántica. Requirió además una corteza cerebral motora y una musculatura más plenamente desarrolladas que nos permitieran realizar sonidos o gestos sofisticados y muy articulados. Con el lenguaje podemos recoger los patrones que aprendemos en la vida y transmitirlos a nuestros hijos y tribu. El lenguaje, sea escrito, hablado o esté incorporado en las tradiciones culturales, se convirtió en el medio con el que transmitimos lo que sabemos del mundo de generación en generación. Hoy la comunicación impresa y electrónica nos permite compartir nuestro conocimiento con millones de personas a lo largo del mundo. Los animales carentes de lenguaje no transmiten tanta información a su progenie. Una rata puede aprender muchos patrones en su vida, pero no lega nueva información detallada: "Mira, hija, así es como me enseñó mi padre a evitar las sacudidas eléctricas".

Así pues, se pueden determinar tres etapas en la inteligencia, empleándose en todas ellas la memoria y la predicción. La primera sería cuando las especies se valían del ADN como medio para establecer la memoria. Los individuos no podían aprender y adaptarse dentro de su vida. Solo eran capaces de transmitir a su descendencia la memoria del mundo basada en el ADN a través de sus genes.

La segunda etapa se inició cuando la Naturaleza inventó sistemas nerviosos modificables, capaces de formar memorias con rapidez. Ahora un individuo lograba aprender cosas importantes sobre la estructura de su mundo y adaptar su conducta en consecuencia durante su vida. Pero todavía no tenía la facultad de comunicar este conocimiento a su prole más que por la observación directa. La creación y expansión de la corteza cerebral ocurrió durante esta segunda etapa, pero no la definió.

La tercera y última etapa es única de los seres humanos. Comienza con la invención del lenguaje y la expansión de nuestra gran corteza cerebral. Nosotros los humanos podemos aprender mucho de la estructura del mundo dentro de nuestras vidas y comunicarlo con efectividad a muchos otros humanos a través del lenguaje. Ustedes y yo estamos participando en este proceso ahora mismo. Yo he pasado gran parte de mi vida indagando en la estructura de los cerebros y en cómo dicha estructura lleva al pensamiento y a la inteligencia. Mediante este libro estoy difundiendo lo que he aprendido. Por supuesto, no podría haberlo hecho si no hubiera tenido acceso al conocimiento reunido por miles de científicos, que aprendieron de otros, y así sucesivamente a través de los siglos. Fui capaz de asimilar y ampliar lo que otros habían escrito sobre su reflexión y observación.



Nos hemos convertido en las criaturas más adaptables del planeta y las únicas con la capacidad de transferir ampliamente nuestro conocimiento del mundo a nuestra especie. La población humana ha pasado por un crecimiento explosivo porque tenemos la facultad de aprender y explotar la estructura del mundo y comunicársela a otros seres humanos. Podemos prosperar en cualquier lugar, sea la selva tropical lluviosa, el desierto o la tundra helada. La combinación de una corteza cerebral grande y el lenguaje ha conducido a la espiral de éxito de nuestra especie.

### **¿Qué es la Creatividad?**

Me preguntan con frecuencia sobre la creatividad, sospecho que porque muchas personas lo consideran algo que una máquina no podría hacer y, por lo tanto, constituye un reto para la idea de construir máquinas inteligentes. ¿Qué es la creatividad? Ya hemos encontrado la respuesta varias veces en este libro. La creatividad no es algo que ocurre en una región particular de la corteza cerebral; tampoco se parece a las emociones o el equilibrio, que tienen su origen en estructuras y circuitos particulares fuera de la corteza cerebral. La creatividad es más bien una propiedad inherente de cada región cortical. Es un componente necesario de la predicción.

¿Cómo puede ser cierto esto? ¿No se trata de una cualidad necesaria que requiere inteligencia y talento elevados? La verdad es que no. La creatividad se puede definir llanamente como la facultad de elaborar predicciones por analogía, algo que ocurre en todas partes de la corteza cerebral y que hacemos de forma continua mientras estamos despiertos. La creatividad ocurre en un continuo. Abarca desde los actos cotidianos sencillos de percepción que suceden en las regiones sensoriales de la corteza cerebral (escuchar una canción en una nueva clave), hasta los actos difíciles y raros de genio que tienen lugar en los niveles superiores de la corteza cerebral (componer una sinfonía en un estilo nuevo). En lo fundamental, los actos cotidianos de percepción son similares a los raros vuelos de la brillantez; lo que pasa es que los actos cotidianos son tan comunes que no les damos importancia.

Ahora ya cuentan con una comprensión básica de cómo creamos memorias invariables, cómo las empleamos para efectuar predicciones y cómo hacemos predicciones de acontecimientos futuros que siempre son algo diferentes de lo que hemos experimentado en el pasado. Recuerden también que nuestras memorias invariables son secuencias de hechos. Efectuamos predicciones combinando el recuerdo de la memoria invariable de lo que deberá pasar a continuación con los detalles pertenecientes a ese momento temporal (recordemos la parábola de la predicción sobre la llegada del tren). La predicción es la aplicación de secuencias de memoria invariable a nuevas situaciones. Por lo tanto, todas las predicciones corticales lo son por analogía. Predecimos el futuro por analogía con el pasado.

Consideremos que estamos a punto de comer en un restaurante desconocido y queremos lavarnos las manos. Aunque jamás hemos estado en ese edificio antes, nuestro cerebro predice que habrá servicios en algún lugar del establecimiento con un lavabo apropiado para lavarse las manos. ¿Cómo se sabe esto? Otros restaurantes en los que hemos estado tienen servicios y por analogía es probable que este también los tenga. Es más, sabemos dónde y qué buscar. Predecimos que habrá una puerta o cartel con algún tipo de símbolo asociado con hombres o mujeres. Predecimos que estará en la parte trasera del restaurante, cerca del bar o de la entrada, pero en general fuera de la vista de las zonas de mesas. De nuevo, nunca hemos estado en este restaurante particular antes, pero por analogía con otros establecimientos similares somos capaces de encontrar lo que necesitamos. No miramos alrededor de forma aleatoria. Buscamos patrones esperados que nos permitan encontrar los servicios enseguida. Este tipo de conducta es un acto creativo; es predecir el futuro por analogía con el pasado. No solemos pensar que esto sea creativo, pero lo es, y mucho.

Hace poco he comprado un vibráfono. Tenemos piano, mas nunca había tocado el vibráfono antes. El día que lo llevamos a casa, cogí una partitura del piano, la coloqué en el atril sobre el vibráfono y comencé a tocar melodías sencillas. Mi habilidad para hacerlo no era notable, pero en lo básico era un acto creativo. Reflexionemos sobre mi actividad. Tengo un instrumento que es muy diferente del piano. El vibráfono posee barras de metal doradas; el piano, teclas negras y blancas. Las barras doradas son grandes y todas del mismo tamaño; las teclas, pequeñas y de dos tamaños diferentes. Las barras doradas están dispuestas en dos filas diferentes; las teclas negras y blancas, intercaladas. En un instrumento empleo los dedos; en el otro, mazos. En un caso, permanezco de pie; en el otro, sentado. Los músculos y movimientos particulares necesarios para tocar el vibráfono son totalmente diferentes de los precisos para tocar el piano.

Así pues, ¿cómo supe tocar una melodía en un instrumento desconocido? La respuesta es que mi corteza cerebral ve una analogía entre las teclas del piano y las barras del vibráfono. El empleo de esta similitud me permitió interpretar la melodía. La verdad es que no es diferente de cantar una canción en una nueva clave. En ambos casos sabemos lo que hacer por analogía con el aprendizaje pasado. Me doy cuenta de que la similitud entre estos dos instrumentos puede parecerles evidente, pero ello se debe a que nuestros cerebros ven analogías de forma automática. Traten de programar un ordenador para que encuentre parecidos entre dos objetos como pianos y vibráfonos, y comprobarán lo difícil que es. La predicción por analogía —la creatividad— es tan dominante que no solemos percibirla.

Sin embargo, sí creemos que somos creativos cuando nuestro sistema de memoria-predicción opera a un nivel de abstracción superior, cuando efectúa predicciones poco comunes utilizando analogías infrecuentes. Por ejemplo, la

mayoría de la gente estaría de acuerdo en que el matemático que demuestra una conjetura difícil es creativo. Pero observemos con detenimiento lo que conllevan sus esfuerzos mentales. Nuestro matemático se concentra mucho en una ecuación y afirma: "¿Cómo voy a desentrañar este problema?". Si la respuesta no es evidente a primera vista, tal vez reordene la ecuación. Al escribirla de forma diferente, puede contemplar el mismo problema desde una perspectiva distinta. Se concentra un poco más. De improviso ve una parte de la ecuación que le resulta conocida. Piensa: "Vaya, reconozco esto. Hay una estructura en esta ecuación que se parece a la estructura de otra ecuación en la que trabajé hace varios años". Entonces efectúa una predicción por analogía: "Tal vez pueda resolver esta ecuación utilizando las mismas técnicas que me sirvieron en el caso de la antigua". Es capaz de resolver el problema por analogía con un problema aprendido previamente. Es un acto creativo.

Mi padre padecía un misterioso trastorno en la sangre que sus médicos no conseguían diagnosticar. ¿Cómo supieron qué tratamiento ofrecer? Una de las cosas que hicieron fue cotejar meses de datos obtenidos del análisis de su sangre por si podían identificar patrones. (Mi padre imprimió un bonito diagrama para que los médicos pudieran ver los datos con claridad.) Aunque sus síntomas no encajaban del todo con los de enfermedades conocidas, había algunas similitudes. Los médicos acabaron basando su tratamiento en una mezcla de estrategias que habían obtenido buenos resultados en otros trastornos sanguíneos. Los tratamientos empleados se dedujeron a partir de analogías con enfermedades que los médicos habían tratado antes. Para reconocer estos patrones se requiere un amplio contacto con otras enfermedades poco comunes.

Las metáforas de Shakespeare son el dechado de la creatividad. "El amor es un humo creado con el vaho de los suspiros"; "filosofía, dulce leche de la adversidad"; "hay dagas en las sonrisas de los hombres". Tales metáforas resultan obvias cuando las vemos, pero son muy difíciles de inventar, razón por la cual se considera a Shakespeare un genio literario. Para crear estas metáforas tuvo que ver una sucesión de analogías inteligentes. Cuando escribe "hay dagas en las sonrisas de los hombres" no habla de dagas o sonrisas. Dagas es análoga a malas intenciones, y las sonrisas de los hombres, a engaño. ¡Dos sabias analogías en solo cinco palabras! Al menos es mi forma de interpretarlas. Los poetas tienen el don de correlacionar palabras o conceptos que no parecen conectados, de forma que dan una nueva luz al mundo. Crean analogías inesperadas como medio para enseñar una estructura de nivel superior.

De hecho, las obras de arte muy creativas son apreciadas porque violan nuestras predicciones. Cuando vemos una película que rompe el molde conocido de un personaje, argumento o cinematografía (incluidos efectos especiales), nos gusta porque no es lo mismo de siempre. La pintura, la música, la poesía y las novelas — todas las formas artísticas creativas— se esfuerzan por romper con lo convencional

y violar las expectativas del público. Existe una tensión contradictoria en lo que hace grande a una obra de arte. Queremos que el arte tenga un aire familiar, pero al mismo tiempo que sea único e inesperado. Si es demasiado conocido, resulta recauchutado o *kitsch*; demasiada singularidad hace que desentone y sea difícil de apreciar. Las obras mejores rompen algunos patrones esperados a la vez que nos enseñan otros nuevos. Observemos una gran obra de música clásica. La mejor música resulta atractiva por su sencillez —buen compás, melodía y fraseo sencillos—. Cualquiera puede entenderla y apreciarla. Sin embargo, también es un poco diferente e inesperada. Pero cuanto más la escuchamos, más vemos que hay un patrón en las partes inesperadas, como armonías poco habituales o cambios de clave repetidos. Lo mismo cabe afirmar de la gran literatura o las grandes películas. Cuanto más las leemos o las vemos, mayor detalle creativo y complejidad de estructura percibimos.

Puede que hayan tenido la experiencia de estar mirando algo y que de repente se les venga a la cabeza una sensación de reconocimiento: “He visto este patrón antes, en algún otro lugar...”. Tal vez no estuvieran intentando resolver un problema, pero una representación invariable de su cerebro se activó por una situación nueva. Han visto una analogía entre dos acontecimientos que no suelen estar relacionados. Quizá yo debiera reconocer que promocionar una idea científica es similar a vender una idea comercial o que propiciar la reforma política es semejante a criar a los hijos. Si fuera poeta, tendría una nueva metáfora. Si fuera científico o ingeniero, tendría una nueva solución para un problema persistente. La creatividad consiste en mezclar y casar patrones de todo lo que hemos experimentado o llegado a conocer en nuestra vida. Es afirmar: “Esto se parece a eso”. El mecanismo neuronal para hacerlo se encuentra en todas partes de la corteza cerebral.

### **¿Son Algunas Personas Más Creativas Que Otras?**

Una pregunta relacionada que suelo escuchar es: “Si todos los cerebros son inherentemente creativos, ¿por qué existen diferencias en nuestra creatividad?”. El marco de la memoria-predicción señala dos respuestas posibles. Una tiene que ver con la Naturaleza, y la otra, con la educación.

En el aspecto educativo, cada cual tiene diferentes experiencias vitales. Por lo tanto, cada cual desarrolla diferentes modelos y memorias del mundo en su corteza cerebral, y efectuaremos analogías y predicciones distintas. Si he tenido contacto con la música, seré capaz de cantar una canción en una nueva clave y tocar melodías sencillas en nuevos instrumentos. Si jamás he tenido contacto con ella, no lograré realizar esos saltos predictivos. Si he estudiado física, seré capaz de explicar la conducta de los objetos cotidianos mediante la analogía con las leyes físicas. Si he crecido con perros, estoy preparado para ver analogías con ellos y lograré predecir mejor sus conductas. Algunas personas son más creativas en situaciones sociales o en lenguaje, matemática o diplomacia según el entorno en el

que crecieron. Nuestras predicciones y, por lo tanto, nuestros talentos se construyen partiendo de nuestras experiencias.

En el capítulo 6 he descrito cómo las memorias se impulsan hacia abajo en la jerarquía cortical. Cuanto más nos exponemos a ciertos patrones, más se reforman las memorias de dichos patrones en niveles inferiores. Esto nos permite aprender la relación entre los objetos abstractos de orden superior que se encuentran en la cima. Es la esencia de la pericia. Un experto es alguien que mediante la práctica y la exposición repetida es capaz de reconocer patrones más sutiles de los que logra determinar alguien no experto, como la forma de una aleta en un coche de finales de la década de 1950 o el tamaño de una mancha en el pico de una gaviota. Los expertos pueden reconocer patrones sobre patrones. El límite físico de lo que somos capaces de reconocer lo marca el tamaño de nuestra corteza cerebral. Pero, como seres humanos, nuestra corteza cerebral es grande comparada con otras especies y gozamos de una tremenda flexibilidad en lo que podemos aprender. Todo depende de las cosas con las que tenemos contacto a lo largo de nuestra vida.

En el aspecto natural, los cerebros muestran una variación física. Sin duda, algunas de las diferencias están determinadas por los genes, como el tamaño de las regiones (los individuos pueden mostrar una diferencia hasta de tres veces en el área bruta V1) y la lateralidad hemisférica (las mujeres tienden a tener un cableado más denso conectando los lados izquierdo y derecho del cerebro que los hombres). Entre los humanos, es probable que algunos cerebros presenten más células o diferentes tipos de conexiones. No parece que el genio creativo de Albert Einstein se diera solo en función del entorno estimulante de la oficina de patentes en la que trabajó de joven. Los análisis recientes de su cerebro —que se creía perdido, pero se encontró conservado en un tarro hace algunos años— revelan que era sensiblemente inusual. Tenía más células de apoyo —llamadas neurogliocitos— por neurona que la media. Mostraba un patrón poco habitual de cisuras o surcos en los lóbulos parietales, región que se considera importante para las capacidades matemáticas y el razonamiento espacial. Además, era un 15 por 100 más amplio que la mayoría de los restantes cerebros. Tal vez nunca sepamos por qué Einstein fue tan creativo e inteligente, pero cabe apostar con total seguridad que parte de su talento se derivó de factores genéticos.

Sea cual fuere la diferencia entre cerebros brillantes y normales, todos somos creativos. Y con la práctica y el estudio podemos aumentar nuestras destrezas y talentos.

### **¿Podemos Entrenarnos para Ser Más Creativos?**

Sí, sin lugar a dudas. He descubierto que hay tres modos de fomentar el hallazgo de analogías útiles cuando se trabaja en la resolución de problemas. Primero es preciso dar por sentado que hay una respuesta a lo que tratamos de resolver. La

gente se rinde con demasiada facilidad. Necesitamos confiar en que hay una solución esperando a ser descubierta y debemos continuar pensando en el problema durante un buen tiempo.

En segundo lugar, tenemos que dejar vagar nuestra mente. Hemos de dar a nuestro cerebro el tiempo y el espacio precisos para descubrir la solución. Encontrar la solución a un problema es, de forma literal, encontrar un patrón almacenado en nuestra corteza cerebral que es análogo al problema en el que estamos trabajando. Si perseveramos en el problema, el modelo de memoria-predicción sugiere que debemos encontrar modos diferentes de considerarlo para aumentar la probabilidad de ver una analogía con una experiencia pasada. Si nos limitamos a sentarnos y mirarlo fijamente una y otra vez, no llegaremos muy lejos. Hemos de tratar de tomar las partes del problema y reordenarlas de forma literal y figurada. Cuando juego al Scrabble, cambio una y otra vez el orden de las fichas. No es que espere que las letras formen por casualidad una nueva palabra, sino que combinaciones diferentes harán que me acuerde de palabras o trozos de palabras que podrían ser parte de una solución. Si miran un dibujo de algo que no tiene sentido, intenten darle la vuelta, cambiar los colores o las perspectivas. Por ejemplo, cuando reflexionaba sobre cómo los diferentes patrones de V1 podrían llevar a representaciones variables en IT, estaba varado, así que di la vuelta al problema para preguntarme cómo un patrón constante de IT era capaz de conducir a predicciones distintas en V1. La inversión del problema resultó útil de inmediato y acabó llevándome a creer que V1 no debía considerarse una región cortical única.

Si se quedan atascados en un problema, abandónenlo por un tiempo. Hagan otra cosa. Luego retómenlo, planteándolo de nuevo. Si lo hacen las veces suficientes, algo surgirá tarde o temprano. Puede llevar días o semanas, pero acabará sucediendo. La meta es encontrar una situación análoga en algún lugar de su experiencia pasada. Para lograrlo deben cavilar sobre el problema a menudo, pero también hacer otras cosas a fin de que la corteza cerebral tenga la oportunidad de encontrar una memoria análoga.

Veamos otro ejemplo de cómo la reordenación de un problema llevó a una solución novedosa. En 1994 mis colegas y yo intentábamos idear cómo introducir texto en ordenadores de mano. Todos nos habíamos centrado en el *software* de reconocimiento de letra. Decíamos: "Escribimos cosas en hojas de papel, así que debemos ser capaces de hacer lo mismo en la pantalla de un ordenador". Por desgracia, resultó ser difícilísimo. Es otra de esas cosas que los ordenadores no hacen nada bien, aunque a los cerebros les resulte muy sencillo. La razón es que el cerebro usa la memoria y el texto en curso para predecir lo que está escrito. Palabras y letras irreconocibles en sí mismas se reconocen con facilidad en el contexto. En el caso de los ordenadores no basta con casar patrones para realizar

la tarea. Yo había diseñado varios ordenadores que empleaban reconocimiento de letra tradicional, pero nunca resultaron demasiado buenos.

Durante varios años me empeñé en lograr que el *software* de reconocimiento funcionara mejor, y estaba atascado. Un día decidí alejarme un poco y contemplar el problema desde una perspectiva diferente. Busqué problemas análogos. Me dije: "¿Cómo introducimos texto en los ordenadores de sobremesa? Lo tecleamos en un teclado. ¿Cómo sabemos cómo teclearlo en un teclado? Bueno, lo cierto es que no es fácil. Es una invención muy reciente y lleva mucho tiempo aprender. Escribir al tacto en un teclado parecido al de una máquina de escribir es difícil y no intuitivo; no se parece en absoluto a escribir, pero millones de personas aprenden a hacerlo. ¿Por qué? Porque funciona". Mi pensamiento continuó por analogía: "Quizá pueda idear un sistema de entrada de texto que no sea necesariamente intuitivo, que se tenga que aprender, pero que la gente utilice porque funciona".

Ese es literalmente el proceso que seguí. Utilicé el acto de escribir en un teclado como analogía para idear cómo meter texto con un puntero en un visualizador. Reconocí que la gente estaba dispuesta a aprender una tarea difícil (teclear) porque era un modo fiable y fácil de introducir texto en una máquina. Por lo tanto, si éramos capaces de crear un nuevo método de introducir texto con un puntero que fuera rápido y fiable, la gente lo utilizaría aunque requiriera aprendizaje. Así pues, diseñé un alfabeto que tradujera de forma fiable lo que escribíamos en texto informático, y lo llamé Graffiti. Con los sistemas de reconocimiento de letra tradicionales, cuando el ordenador malinterpreta nuestra escritura no sabemos por qué. Pero el sistema Graffiti siempre produce la letra correcta a menos que cometamos un error al escribir. Nuestros cerebros aborrecen lo impredecible, motivo por el cual la gente siente aversión por los sistemas de reconocimiento de letra tradicionales.

Mucha gente pensó que Graffiti era una idea completamente estúpida. Iba en contra de todo lo que se creía sobre cómo se suponía que funcionaban los ordenadores. En esos días el mantra era que los ordenadores tenían que adaptarse al usuario y no al contrario. Pero yo confiaba en que la gente aceptaría este nuevo modo de introducir texto por analogía con el teclado. Graffiti resultó una buena solución y su adopción fue amplia. Hasta la fecha sigo oyendo a la gente declarar que los ordenadores deben adaptarse a los usuarios, pero no siempre es verdad. Nuestros cerebros prefieren sistemas coherentes y predecibles, y nos gusta aprender nuevas destrezas.

### **¿Puede la Creatividad Descarriarme? ¿Puedo Engañarme a Mí Mismo?**

La falsa analogía es siempre un peligro. La historia de la ciencia está repleta de bellas analogías que resultaron erróneas. Por ejemplo, el famoso astrónomo Johannes Kepler llegó al convencimiento de que las órbitas de los seis planetas conocidos estaban definidas por los sólidos platónicos. Estos son las únicas formas

tridimensionales que pueden construirse por completo partiendo de polígonos regulares. Existen solo cinco: tetraedro (cuatro triángulos equiláteros), hexaedro (seis cuadrados, también conocido como cubo), octaedro (ocho triángulos equiláteros), dodecaedro (doce pentágonos regulares) e icosaedro (veinte triángulos equiláteros). Fueron descubiertos por los griegos antiguos, que estaban obsesionados con la relación de la matemática y el Cosmos.

Al igual que todos los eruditos renacentistas, Kepler estaba muy influido por el pensamiento griego. Le parecía que no podía ser una coincidencia que hubiera cinco sólidos platónicos y seis planetas. Así lo expresa en su libro *El misterio del Cosmos* (1596): "El mundo dinámico está representado por los sólidos de caras planas. Hay cinco, pero cuando se los considera límites, esos cinco determinan seis cosas distintas: de ahí los seis planetas que giran alrededor del Sol. También es esta la razón por la que no hay más que seis planetas". Vio una analogía hermosa, pero completamente falsa.

Kepler proseguía explicando las órbitas de los planetas recurriendo a la jerarquización de los sólidos platónicos con el Sol en el centro. Tomó la esfera definida por la órbita de Mercurio como línea de referencia y circunscribió en ella un octaedro cuyas puntas definían una esfera mayor, que proporcionaba la órbita de Venus. Alrededor de esta circunscribió un icosaedro, cuyas puntas exteriores producían la órbita de la Tierra. La progresión continuaba: el dodecaedro dibujado alrededor de la órbita terrestre daba la órbita de Marte; el tetraedro en torno a esta señalaba la órbita de Júpiter, y el cubo alrededor de esta indicaba la órbita de Saturno. Era elegante y bello. Dada la precisión limitada de los datos astronómicos de su época, pudo convencerse de que este esquema funcionaba. (Años más tarde, Kepler se dio cuenta de que se había equivocado tras apoderarse de los precisos datos astronómicos de su colega fallecido Tycho Brahe, que demostraban que las órbitas planetarias son elipses y no círculos.)

La exaltación de Kepler sirve como moraleja para los científicos y, en realidad, para todos los pensadores. El cerebro es un órgano que construye modelos y efectúa predicciones creativas, pero sus modelos y predicciones pueden ser tanto engañosos como válidos. Nuestros cerebros siempre están buscando patrones y haciendo analogías. Si no se pueden encontrar correlaciones acertadas, el cerebro se muestra más que contento de aceptar las falsas. La seudociencia, el fanatismo, la fe y la intolerancia suelen tener sus raíces en la falsa analogía.

### **¿Qué es la Conciencia?**

Es una de esas preguntas que temen los neurocientíficos, en mi opinión de forma innecesaria. Algunos científicos, como Christof Koch, están dispuestos a desentrañar el tema de la conciencia, pero la mayoría lo consideran un asunto de la filosofía que bordea la seudociencia. Pienso que merece consideración aunque solo sea porque despierta la curiosidad de mucha gente. No puedo proporcionar



una respuesta completamente satisfactoria, mas creo que la memoria y la predicción la abordan en parte. En primer lugar, hablemos de ese acertijo tal como surge en la conversación.

No hace mucho tiempo me encontraba en una conferencia científica en un bonito lugar de Long Island Sound. A primeras horas de la tarde, una docena de asistentes cogimos nuestros vasos de vino y los llevamos hasta un embarcadero para sentarnos junto al agua y charlar antes de la comida y la sesión vespertina. Transcurrido un rato, la conversación giró al asunto de la conciencia. Como ya he dicho, los neurocientíficos no suelen hablar de eso, pero estábamos en un bonito paraje, habíamos bebido vino y surgió el tema.

Una científica británica soltó una perorata sobre sus ideas al respecto y afirmó: "Por supuesto, jamás comprenderemos la conciencia". Yo no estaba de acuerdo: "La conciencia no es un gran problema. Creo que no es más que lo que se siente al tener corteza cerebral". Se hizo un silencio en el grupo y luego se suscitó una discusión cuando varios científicos intentaron ilustrarme sobre mi error evidente. "Debe admitir que el mundo parece vivo y hermoso. ¿Cómo puede negar que su conciencia percibe el mundo? Debe admitir que se siente como algo especial." Para establecer mi premisa, repliqué: "No sé de qué hablan. Dado el modo en que se están expresando sobre la conciencia, tengo que concluir que soy diferente de ustedes. Yo no siento lo que ustedes están sintiendo, así que tal vez no sea un ser consciente. Debo de ser un zombi". Se suele recurrir a los zombis cuando los filósofos hablan sobre la conciencia. Un zombi se define como alguien físicamente igual que un humano, mas sin conciencia. Son máquinas de carne y hueso que caminan y respiran, pero sin nadie dentro.

La científica británica me miró. "Claro que es usted consciente."

"No, no lo creo. Puede que se lo parezca, pero no soy un ser humano consciente. No se preocupe por ello; me encuentro a gusto."

Ella repuso, "Bueno, ¿es que no percibe la maravilla?" y extendió el brazo hacia el agua centelleante mientras el sol comenzaba a hundirse y el cielo se volvía rosa salmón iridiscente.

"Sí, veo todas esas cosas; ¿y qué?"

"Entonces, ¿cómo explica su experiencia subjetiva?"

Repliqué, "Sí, sé que estoy aquí. Tengo memoria de cosas semejantes a este atardecer. Pero no me parece que esté sucediendo nada especial, así que si usted siente algo especial quizá es que yo no soy consciente." Intentaba hacerle definir lo que pensaba que había de maravilloso e inexplicable en la conciencia. Trataba de lograr que la definiera.

Continuamos en esta línea de razonamiento —sí, lo es; no, yo no soy eso— hasta que llegó el momento de dirigirnos a comer. No creo que cambiara el modo de pensar de nadie sobre la existencia y el significado de la conciencia. Pero trataba de lograr que se dieran cuenta de que la mayoría de la gente piensa que es una especie de salsa mágica que se añade sobre el cerebro físico. Tenemos un cerebro, compuesto de células, y vertemos la conciencia, esa salsa mágica, sobre él, y esa es la condición humana. En este planteamiento, la conciencia es una entidad misteriosa separada de los cerebros. Por eso los zombis tienen cerebros, pero no conciencia. Poseen todos los elementos mecánicos, neuronas y sinapsis, mas carecen de la salsa especial. Pueden hacer todo lo que un humano hace. Por el exterior no se puede distinguir a un zombi de un humano.

La idea de que la conciencia es algo añadido proviene de creencias anteriores en el *élan vital*, una fuerza especial que se pensaba que animaba a los seres vivos. La gente creía que se necesitaba dicha fuerza para explicar la diferencia entre las rocas y las plantas o los metales y las doncellas. Pocos lo siguen creyendo. En la actualidad sabemos bastante sobre las diferencias entre materia inanimada y animada para entender que no hay una salsa especial. Ahora sabemos mucho sobre ADN, plegamiento de proteínas, transcripción de genes y metabolismo. Aunque todavía no conocemos todos los mecanismos de los sistemas vivos, sí sabemos lo suficiente de biología para abandonar la magia. De forma similar, la gente ya no sugiere que se necesita magia o espíritus para hacer que los músculos se muevan. Tenemos proteínas que se pliegan y forman grandes cadenas de moléculas. Se puede leer todo al respecto.

No obstante, mucha gente persiste en creer que la conciencia es diferente y no puede explicarse en términos biológicos reduccionistas. Reitero que no soy un estudioso de la conciencia. No he leído todas las opiniones de los filósofos, pero tengo algunas ideas sobre lo que confunde a la gente en este debate. Creo que la conciencia es lo que se siente al tener corteza cerebral. Pero lo podemos mejorar. Podemos dividir la conciencia en dos categorías importantes. Una es similar al conocimiento de uno mismo, la noción cotidiana de estar consciente, y es relativamente fácil de entender. La segunda son los *qualia*, la idea de que los sentimientos asociados con la sensación son en cierto modo independientes de la entrada sensorial. Es la parte más difícil.

Cuando la mayoría de la gente dice la palabra *consciente* se está refiriendo a la primera categoría. “¿Fuiste consciente de que pasaste a mi lado sin decirme hola?” “¿Eras consciente cuando te caíste de la cama la noche pasada?” “No eres consciente cuando duermes.” Algunas personas afirman que esta forma de conciencia es lo mismo que el conocimiento. Están muy próximos en significado, pero no creo que su conocimiento la capte con acierto. Sugiero que este significado de conciencia es sinónimo de formar memorias declarativas. Son memorias que se pueden recordar y hablar de ellas a otras personas. Se pueden

expresar verbalmente. Si me preguntan dónde fui el fin de semana pasado, puedo decírselo. Es una memoria declarativa. Si me preguntan cómo equilibrar una bicicleta, puedo decirles que sostengan la barra del manillar y empujen los pedales, pero soy incapaz de explicarles con exactitud cómo hacerlo. El modo de equilibrar una bicicleta tiene mucho que ver con la actividad del cerebro antiguo, así que no es una memoria declarativa.

Cuento con un pequeño experimento mental para mostrar que nuestra noción cotidiana de conciencia se corresponde con formar memorias declarativas. Recordemos que se piensa que todas las memorias residen en los cambios físicos de las sinapsis y las neuronas que conectan. Por lo tanto, si tuviera un método para invertir esos cambios físicos, nuestras memorias se borrarían. Ahora imaginemos que pudiera instalar un interruptor y devolver su cerebro al estado físico exacto en que se hallaba en algún momento del pasado. Podría ser hace una hora, veinticuatro horas o cualquier otra cosa. Yo solo pulso el interruptor en mi máquina de vuelta al pasado y sus sinapsis y neuronas regresan a un estado previo en el tiempo. Al hacerlo, borro toda su memoria de lo que ha ocurrido desde entonces.

Supongamos que viven el día de hoy y se despiertan mañana. Pero en el momento en que están despertándose, pulso el interruptor y borro las últimas veinticuatro horas. Tendrían una memoria cero del día anterior. Desde la perspectiva de sus cerebros, ayer nunca sucedió. Les diría que es miércoles y ustedes protestarían: "No, es martes. Estoy seguro. Han cambiado el calendario. Es martes, sin duda. ¿Por qué me está gastando esta broma?". Pero todas las personas con las que se encontraron el martes dirían que habían estado conscientes durante todo el día. Los vieron, comieron y hablaron con ustedes. ¿No lo recuerdan? Responden que no; no ha sucedido. Por último, cuando se les muestra un vídeo en el que se los ve comiendo, van convenciéndose poco a poco de que existió ese día, aunque no tengan memoria de él. Es como si hubieran sido zombis durante un día, sin conciencia. Sin embargo, fueron conscientes todo el tiempo. Su creencia de que eran conscientes solo desapareció cuando su memoria declarativa se borró.

Este experimento mental capta la equivalencia entre memoria declarativa y nuestra noción cotidiana de ser conscientes. Si durante un partido de tenis a su término les pregunto si son conscientes, me contestarían sin duda que sí. Si luego borro su memoria de las dos últimas horas, declararían que estaban inconscientes y no eran responsables de sus acciones durante ese tiempo. En ambos casos jugaron el mismo partido de tenis. La única diferencia es si tienen memoria de ello en el momento en que se lo pregunto. Por lo tanto, este significado de conciencia no es absoluto. Puede cambiarse tras el hecho mediante el borrado de la memoria.

La cuestión más difícil sobre la conciencia atañe a los *qualia*, que se suelen formular en preguntas de tipo zen, como: "¿Por qué el rojo es rojo, y el verde, verde? ¿Me parece el rojo igual que a ti? ¿Por qué el rojo está cargado

emocionalmente de ciertos sentimientos? Sin duda, posee para mí una cualidad o sensación inextricable. ¿Qué sentimientos te causa a ti?”.

Encuentro dichas descripciones difíciles de relacionar con la neurobiología, así que me gustaría replantear la pregunta. A mi entender, una pregunta equivalente, aunque también difícil de explicar, sería por qué los diversos sentidos parecen cualitativamente diferentes. ¿Por qué la vista parece diferente del oído y por qué el oído parece diferente del tacto? Si la corteza cerebral es la misma en todas partes, si funciona con los mismos procesos, si se limita a manejar patrones, si no entra ningún sonido o luz en el cerebro, solo patrones, ¿por qué la visión parece tan diferente del oído? Me resulta difícil describir en qué difiere la vista del oído, pero es algo evidente. Supongo que también lo es para ustedes. No obstante, el axón que representa el sonido y otro que representa la luz son idénticos a todos los efectos prácticos. Las cualidades de la luz y del sonido no se transportan en el axón de una neurona sensorial.

La gente que presenta una afección llamada sinestesia tiene cerebros que velan la distinción entre los sentidos: algunos sonidos o ciertas texturas tienen color. Esto nos indica que el aspecto cualitativo de un sentido no es inmutable. Mediante algún tipo de modificación física, un cerebro puede conceder un aspecto cualitativo de visión a una entrada auditiva.

Así pues, ¿cuál es la explicación para los *qualia*? Puedo pensar en dos posibilidades, pero ninguna de ellas me resulta totalmente satisfactoria. Una es que aunque el oído, el tacto y la visión funcionan según principios similares de la corteza cerebral, se manejan de forma diferente por debajo de esta. El oído descansa en un conjunto de estructuras subcorticales específicas para la audición que procesan los patrones auditivos antes de que lleguen a la corteza cerebral. Los patrones somatosensoriales también viajan por un conjunto de áreas subcorticales que son únicas para los sentidos somáticos. Tal vez los *qualia*, como las emociones, no estén mediatizados simplemente por la corteza cerebral. Si están vinculados de algún modo con partes subcorticales del cerebro que poseen una conexión única, tal vez ligada con los centros de emoción, ello podría explicar por qué los percibimos de forma diferente, aunque no ayude a resolver por qué existe el tipo de sensación de *qualia*.

La otra posibilidad que se me ocurre es que la estructura de las entradas —diferencias entre los mismos patrones— dicte cómo experimentamos los aspectos cualitativos de la información. La naturaleza del patrón espacial-temporal en el nervio auditivo es diferente de la naturaleza del mismo patrón en el nervio óptico. Este último posee millones de fibras y transporta bastante información espacial. El nervio auditivo solo tiene treinta mil fibras y transporta más información temporal. Estas diferencias puede que estén relacionadas con lo que denominamos *qualia*.

De lo que no cabe duda alguna es de que, se defina como se defina la conciencia, la memoria y la predicción desempeñan un papel crucial en su creación.

Las nociones de mente y alma están relacionadas con la conciencia.

De niño solía preguntarme cómo habría sido si "yo" hubiera nacido en el cuerpo de otro niño en otro país, como si "yo" fuera algo independiente de mi cuerpo. Este sentimiento de una mente independiente de la fisicalidad es común y una consecuencia natural del funcionamiento de la corteza cerebral. Nuestra corteza cerebral crea un modelo del mundo en su memoria jerárquica. Los pensamientos son lo que surge cuando este modelo funciona por sí mismo; el recuerdo de la memoria lleva a un nuevo recuerdo de la memoria, y así sucesivamente. Nuestros pensamientos más contemplativos no están dirigidos por el mundo real, ni siquiera están conectados con él; son puras creaciones de nuestro modelo. Cerramos los ojos y buscamos en silencio para que nuestro pensamiento no se vea interrumpido por las entradas sensoriales. Nuestro modelo, por supuesto, se creó originalmente mediante la exposición al mundo real a través de los sentidos, pero cuando planeamos y reflexionamos sobre él lo hacemos mediante el modelo cortical, no el mundo en sí.

Para la corteza cerebral, nuestros cuerpos no son más que parte del mundo exterior. Recordemos que el cerebro es una caja silenciosa y oscura. Solo sabe del mundo a través de los patrones de las fibras nerviosas sensoriales. Desde la perspectiva del cerebro como un aparato de patrones, no sabe de nuestro cuerpo de forma diferente de como sabe del resto del mundo. No existe una distinción especial entre dónde termina el cuerpo y dónde empieza el resto del mundo. Pero la corteza cerebral no tiene capacidad para modelar al cerebro porque no hay sentidos en él. De este modo, podemos ver por qué nuestros pensamientos parecen independientes de nuestro cuerpo, por qué parece que tenemos una mente o alma independiente. La corteza cerebral construye un modelo de nuestro cuerpo, pero no puede construir un modelo del mismo cerebro. Nuestros pensamientos, que se localizan en el cerebro, están separados físicamente del cuerpo y el resto del mundo. La mente es independiente del cuerpo, pero no del cerebro.

Podemos observar con claridad esta diferenciación mediante el trauma y la enfermedad. Si alguien pierde un miembro, el modelo que tiene su cerebro de él puede permanecer intacto, dando como resultado el denominado "miembro fantasma" que es capaz de seguir sintiendo unido a su cuerpo. Por otro lado, si sufre un trauma cortical, puede perder su modelo del brazo aunque siga conservándolo. En este caso puede padecer el síndrome del brazo ajeno y tener la sensación incómoda y quizá intolerable de que el brazo no es el suyo y es controlado por otra persona. Si nuestro cerebro permanece intacto mientras el resto del cuerpo cae enfermo, nos da la sensación de que tenemos una mente sana atrapada en un cuerpo agonizante, aunque la realidad es que tenemos un

cerebro sano atrapado en un cuerpo agonizante. Es natural imaginar que nuestra mente continuará después de la muerte de nuestro cuerpo, pero cuando el cerebro muere, también lo hace la mente. Esta verdad resulta evidente si nuestros cerebros fallan antes que nuestros cuerpos. La gente con la enfermedad de Alzheimer o con un daño cerebral serio pierde la mente aunque su cuerpo permanezca sano.

### **¿Qué es la Imaginación?**

Desde el punto de vista conceptual, la imaginación es bastante simple. Los patrones fluyen a cada área cortical desde nuestros sentidos o áreas inferiores de la jerarquía de la memoria. Cada área cortical crea predicciones que se reenvían jerarquía abajo. Para imaginar algo, basta con dejar que nuestras predicciones den la vuelta y se conviertan en entradas. Sin hacer nada físico, podemos seguir las consecuencias de nuestras predicciones. "Si sucede esto, sucederá eso y luego aquello", y así sucesivamente. Lo experimentamos cuando preparamos una reunión de negocios, jugamos al ajedrez, organizamos un acontecimiento deportivo o hacemos miles de otras cosas.

En el ajedrez nos imaginamos que movemos el caballo a cierta posición y luego imaginamos cómo quedará el tablero tras la jugada. Con esta imagen en mente, predecimos lo que hará nuestro rival y cómo quedará el tablero después de su jugada. Luego predecimos lo que haremos, y así una y otra vez. Avanzamos por los pasos imaginados y sus consecuencias. Al final decidimos, basándonos en estas secuencias de hechos imaginados, si la jugada inicial era buena o no. Algunos atletas, como los esquiadores de descenso contrarreloj, pueden mejorar sus resultados si repasan mentalmente el recorrido de la carrera una y otra vez en su cabeza. Al cerrar los ojos e imaginarse cada uno de los giros y los obstáculos, e incluso encontrarse en el podio de los ganadores, aumentan sus posibilidades de éxito. Imaginar no es más que otra palabra para planificar. Ahí es donde la capacidad predictiva de nuestra corteza cerebral merece la pena, pues nos permite saber cuáles serán las consecuencias de nuestras acciones antes de que las realicemos.

Imaginar requiere un mecanismo neuronal para convertir una predicción en una entrada. En el capítulo 6 propuse que en las células de la capa 6 es donde ocurre la predicción precisa. Estas células se proyectan a niveles inferiores de la jerarquía, pero también hacia arriba a las células de entrada de la capa 4. De este modo, las salidas de una región pueden convertirse en sus propias entradas. Stephen Grossberg, que lleva mucho tiempo realizando modelos de la corteza cerebral, denomina a este circuito de la imaginación "realimentación plegada". Si cerramos los ojos e imaginamos un hipopótamo, el área visual de nuestra corteza cerebral se activa, igual que lo haría si estuviéramos viendo realmente el animal. Vemos lo que imaginamos.

## ¿Qué es la Realidad?

La gente me pregunta con una expresión de preocupación y asombro, "¿Quiere decir que nuestros cerebros crean un modelo del mundo? ¿Y que el modelo puede ser más importante que la realidad?"

"Bueno, sí; en cierta medida, así sería," respondo.

"¿Pero no existe el mundo fuera de mi cabeza?"

Por supuesto que sí. La gente es real, mi gata es real, las situaciones sociales en las que nos encontramos son reales. Pero nuestra comprensión del mundo y nuestra respuesta a ella se basan en predicciones que provienen de nuestro modelo interno. En un momento dado del tiempo, solo podemos sentir directamente una parte diminuta de nuestro mundo. Esa ínfima parte dicta qué memorias se invocarán, pero no basta en sí misma para construir el conjunto de nuestra percepción en curso. Por ejemplo, ahora estoy tecleando en mi despacho y escucho una llamada en la puerta delantera. Sé que mi madre ha venido de visita e imagino que está al pie de la escalera, aunque no la he visto ni escuchado en realidad. No había nada en la entrada sensorial ligado de forma específica con mi madre. Es mi modelo de memoria del mundo el que predice que está ahí por analogía con la experiencia pasada. La mayoría de lo que percibimos no llega a través de nuestros sentidos, sino que es generado por nuestro modelo de memoria interno.

Así pues, la pregunta "¿Qué es la realidad?" depende en buena medida de la precisión con la que nuestro modelo cortical refleje la verdadera naturaleza del mundo.

Muchos aspectos del mundo que nos rodea son tan constantes que casi todos los humanos poseen el mismo modelo interno de ellos. De niños aprendimos que la luz que cae sobre un objeto redondo produce determinada sombra y que podemos calcular la forma de la mayoría de los objetos por las pistas del mundo natural. Aprendimos que si lanzábamos una taza desde nuestra trona, la gravedad siempre la empujaba al suelo. Aprendimos texturas, geometría, colores y los ritmos del día y la noche. Todos aprendemos de forma sistemática las propiedades físicas sencillas del mundo.

Pero buena parte de nuestro modelo del mundo se basa en la costumbre, la cultura y lo que nuestros padres nos enseñan. Estas partes de nuestro modelo son menos constantes y podrían ser muy diferentes para distintas personas. Un niño que se cría en un hogar lleno de cariño y cuidados con padres que responden a sus necesidades emocionales es probable que llegue a la edad adulta prediciendo que el mundo es seguro y amable. Los niños maltratados por uno o ambos padres tienen mayores probabilidades de ver los acontecimientos futuros como algo

peligroso y cruel, y creer que no se puede confiar en nadie por muy bien que se los trate después. Gran parte de la psicología se basa en las consecuencias de la experiencia, el apego y la educación durante los primeros años, porque es cuando el cerebro establece su primer modelo del mundo.

Nuestra cultura moldea por completo nuestro modelo del mundo. Por ejemplo, estudia cómo los asiáticos y los occidentales perciben el espacio y los objetos de forma diferente. Los asiáticos atienden más al espacio entre los objetos, mientras que los occidentales se fijan en los objetos, una diferencia que se traduce en estéticas y modos de resolver problemas separados. La investigación ha demostrado que algunas culturas, como en el caso de las tribus de Afganistán y algunas comunidades de América del Sur, se construyen sobre los principios del honor y, como resultado, son más proclives a aceptar el carácter natural de la violencia. Los credos religiosos diferentes aprendidos al comienzo de la vida pueden llevar a modelos completamente distintos de moralidad, modos de tratar a los hombres y las mujeres e incluso del valor de la vida misma. Es evidente que estos modelos del mundo diferentes no pueden ser todos acertados en un sentido absoluto y universal, aunque tal vez parezcan correctos a un individuo. El razonamiento moral, lo bueno y lo malo, se aprende.

Nuestra cultura (y la experiencia familiar) nos enseña estereotipos, que por desgracia son una parte inevitable de la vida. A lo largo de este libro se podría sustituir *memoria invariable* (o representación invariable) por la palabra *estereotipo* sin que se alterara de forma sustancial el significado. La predicción por analogía es muy semejante al juicio por estereotipo. El estereotipo negativo tiene consecuencias sociales terribles. Si mi teoría de la inteligencia es acertada, no podemos lograr que la gente se libre de su propensión a pensar en estereotipos, porque son el modo de funcionar de la corteza cerebral. Formar estereotipos es un rasgo inherente del cerebro.

El modo de eliminar el daño causado por los estereotipos es enseñar a nuestros hijos a reconocer los falsos, a ser empáticos y escépticos. Necesitamos fomentar estas capacidades de pensamiento crítico, además de inculcarles los mejores valores que conocemos. El escepticismo, el núcleo del método científico, es el único modo que entendemos para separar el hecho de la ficción.

\* \* \*

**E**spero haberles convencido ya de que la mente no es más que una etiqueta de lo que hace el cerebro. No es una cosa separada que manipula o coexiste con las células en el cerebro. Las neuronas no son más que células. No existe una fuerza mística que haga que las células nerviosas particulares o los grupos de células nerviosas se comporten de maneras que difieren de como lo harían normalmente. Debido a este hecho, podemos ahora girar nuestra atención al modo como



podríamos aplicar la capacidad de las células cerebrales para recordar y predecir — nuestro algoritmo cortical— en silicio.

## El Futuro de la Inteligencia

**Es** difícil predecir los usos finales de una nueva tecnología. Como hemos visto a lo largo de este libro, los cerebros efectúan predicciones por analogía con el pasado. Así pues, nuestra inclinación natural es imaginar que una nueva tecnología se empleará para hacer el mismo tipo de cosas que efectuaba otra anterior. Imaginamos la utilización de una nueva herramienta para realizar algo conocido, solo que más deprisa, con mayor eficacia o más barato.

Abundan los ejemplos. La gente llamó al ferrocarril el “caballo de hierro”, y al automóvil, el “carruaje sin caballos”. Durante décadas, el teléfono se consideró dentro del contexto del telégrafo, algo que solo debía usarse para comunicar noticias o sucesos importantes; hasta la década de 1920 la gente no empezó a emplearlo de manera informal. La fotografía se usó al principio como una nueva forma de retrato. Y las películas cinematográficas se conceptuaron como una variación de las obras de teatro, motivo por el cual las salas de cine tuvieron telones que se corrían sobre la pantalla durante buena parte del siglo XX.

No obstante, los usos finales de una nueva tecnología son a menudo inesperados y de mucho mayor alcance de lo que nuestra imaginación puede captar al principio. El teléfono ha evolucionado a una red inalámbrica de voz y datos que permite a dos personas cualesquiera del planeta comunicarse entre sí estén donde estén mediante voz, texto e imágenes. El transistor fue inventado por Bell Labs en 1947. De inmediato resultó evidente que el aparato era un gran avance, pero las aplicaciones iniciales no fueron más que mejoras de otras antiguas: los transistores reemplazaron a los tubos de vacío. Esto llevó a la fabricación de radios y ordenadores menores y más fiables, que fue algo importante y apasionante en su época, pero las principales diferencias estribaban en el tamaño y fiabilidad de las máquinas. Las aplicaciones más revolucionarias de los transistores no se descubrieron hasta más tarde. Fue necesario un período de innovación gradual antes de que alguien pudiera concebir el circuito integrado, el microprocesador, el procesador de señales digitales o el chip de memoria. Asimismo, el microprocesador fue desarrollado por primera vez en 1970 teniendo en mente las calculadoras de sobremesa. Una vez más, las primeras aplicaciones se limitaron a sustituir tecnologías existentes. La calculadora electrónica reemplazó a la calculadora de sobremesa mecánica. Los microprocesadores también fueron candidatos claros para reemplazar a los solenoides que entonces se empleaban en ciertos tipos de control industrial, como los cambios de los semáforos. Sin embargo, años después empezó a manifestarse su verdadero poder. Nadie de esa época fue capaz de prever el ordenador personal moderno, el teléfono móvil, Internet, el *Global Positioning System* (GPS) o cualquier otra de las piezas básicas de la tecnología de la información actual.

De igual manera, sería absurdo pensar que podemos predecir las aplicaciones revolucionarias de los sistemas de memoria semejantes al cerebro. Espero que estas máquinas mejoren la vida en todos los sentidos. Podemos estar seguros de que así será. Pero predecir el futuro de la tecnología más allá de unos cuantos años es imposible. Para darse cuenta de ello no hay más que leer algunos de los absurdos pronósticos que han realizado los futuristas a lo largo de los años. En la década de 1950 se predijo que para el año 2000 todos tendríamos reactores atómicos en nuestros sótanos y pasaríamos las vacaciones en la Luna. Pero mientras no olvidemos la moraleja de estos cuentos, no se pierde nada por especular cómo serán las máquinas inteligentes. Como mínimo, se pueden extraer sobre el futuro algunas conclusiones generales y útiles.

Las preguntas resultan intrigantes. ¿Podemos construir máquinas inteligentes y, si es así, qué apariencia tendrán? ¿Se aproximarán a los robots parecidos a los humanos de la ficción popular, a la caja negra o beis de un ordenador personal, o a alguna otra cosa? ¿Cómo se usarán? ¿Se trata de una tecnología peligrosa que puede dañarnos o amenazar nuestras libertades personales? ¿Cuáles son las aplicaciones obvias de las máquinas inteligentes y hay algún modo de saber cuáles serán las aplicaciones fantásticas? ¿Cuál será la repercusión final de las máquinas inteligentes sobre nuestras vidas?

### **¿Podemos Construir Máquinas Inteligentes?**

Sí podemos, pero quizá no sean lo que esperamos. Aunque tal vez parezca lo obvio, no creo que vayamos a construir máquinas inteligentes que actúen como seres humanos, ni interactúen con nosotros de formas humanas.

Una noción popular de las máquinas inteligentes nos llega de las películas y libros: son los robots humanoides amables, malos o a veces torpes que conversan con nosotros sobre sentimientos, ideas y acontecimientos, y desempeñan un papel en innumerables argumentos fantásticos. Un siglo de ciencia ficción ha entrenado a la gente para que considere a los robots y andróides una parte inevitable y deseable de nuestro futuro. Han crecido generaciones con las imágenes de Robbie, el robot de *Planeta prohibido*; R2D2 y C3PO, de *La guerra de las galaxias*, y el teniente comandante Data, de *Star Trek*. Incluso HAL, de la película *2001: Una odisea del espacio*, aunque carente de cuerpo, era muy parecido a los humanos y había sido diseñado por estos tanto para ser un compañero como un copiloto programado en su largo viaje espacial. Los robots de aplicaciones limitadas —cosas como coches inteligentes, minisubmarinos autónomos para explorar las profundidades oceánicas y aspiradoras o cortacéspedes auto-dirigidos— son factibles, y es muy posible que algún día sean de uso corriente. Pero andróides y robots como el teniente comandante Data y C3PO van a seguir siendo ficción durante mucho tiempo. Hay un par de razones para ello.

La primera es que la mente humana no solo la crea la corteza cerebral, sino los sistemas emocionales del cerebro viejo y la complejidad del cuerpo humano. Para ser humano se necesita toda la maquinaria biológica, no solo la corteza cerebral. Para conversar como un ser humano sobre todos los temas (para pasar el test de Turing) se requeriría una máquina inteligente que tuviera la mayoría de las experiencias y emociones de un humano real y viviera una vida semejante a la humana. Las máquinas inteligentes contarán con el equivalente de una corteza cerebral y un conjunto de sentidos, pero el resto es optativo. Tal vez resulte entretenido observar a una máquina inteligente desplazarse por ahí en cuerpos parecidos a los humanos, mas no poseerán mentes que sean remotamente semejantes a las humanas a menos que se les incorporen sistemas emocionales y experiencias semejantes a las nuestras, lo cual sería difícilísimo y, a mi parecer, carecería de sentido.

La segunda razón es que, dado el coste y el esfuerzo necesarios para construir y mantener robots humanoides, cuesta ver que resultaran prácticos. Un mayordomo robot sería más caro y menos útil que un asistente humano. Aunque el robot fuera "inteligente", no tendría la relación y comprensión fácil que sí muestra un asistente humano por el simple hecho de ser un semejante humano.

Tanto la máquina de vapor como el ordenador digital suscitaron visiones robóticas que nunca llegaron a fructificar. De igual modo, cuando pensamos en construir máquinas inteligentes, a muchas personas les resulta natural imaginar robots humanoides, pero no es probable que se haga realidad. Los robots son un concepto nacido de la Revolución Industrial y perfeccionado por la ficción. No debemos acudir a ellos en busca de inspiración para desarrollar máquinas verdaderamente inteligentes.

Así pues, ¿qué apariencia tendrán las máquinas inteligentes si no son robots que caminan y hablan? La evolución descubrió que si unía a nuestros sentidos un sistema de memoria jerárquico, la memoria modelaría el mundo y predeciría el futuro. Copiando a la Naturaleza, debemos construir máquinas inteligentes con los mismos principios. Ahí está la receta para construir dichas máquinas. Se comienza con un conjunto de sentidos para extraer patrones del mundo. Nuestra máquina inteligente puede que tenga un conjunto de sentidos diferentes de los humanos e incluso tal vez "exista" en un mundo distinto al nuestro. Por lo tanto, no hay que dar por sentado que ha de tener un juego de globos oculares y un par de orejas. A continuación se debe añadir a esos sentidos un sistema de memoria jerárquico que funcione según los mismos principios que la corteza cerebral. Luego tendremos que entrenar a ese sistema de memoria de forma muy semejante a como se enseña a los niños. Tras sesiones de entrenamiento repetitivo, nuestra máquina inteligente construirá un modelo de *su* mundo tal como lo ven *sus* sentidos. No habrá necesidad ni oportunidad para que nadie le programe las reglas del mundo, bases de datos, hechos o cualesquiera de los conceptos de alto nivel que son la

pesadilla de la inteligencia artificial. La máquina inteligente debe aprender a través de la observación de su mundo e incluir las entradas de un instructor cuando sea preciso. Una vez que nuestra máquina inteligente ha creado un modelo de su mundo, puede descubrir analogías con experiencias pasadas, efectuar predicciones sobre acontecimientos futuros, proponer soluciones a nuevos problemas y poner a nuestra disposición dicho conocimiento.

Desde el punto de vista físico, nuestra máquina inteligente puede incorporarse a aviones o coches, o permanecer estoicamente en un estante de una sala de ordenadores. A diferencia de los seres humanos, cuyos cerebros deben acompañar a sus cuerpos, el sistema de memoria de una máquina inteligente podría emplazarse muy lejos de sus sensores (y "cuerpo", si lo tuviera). Por ejemplo, un sistema de seguridad inteligente podría tener sensores colocados por toda una fábrica o ciudad, pero el sistema de memoria jerárquico unido a esos sensores estaría encerrado en el sótano de un edificio. Por lo tanto, la encarnación física de una máquina inteligente tendría la posibilidad de adoptar muchas formas.

No hay razón para que una máquina inteligente presente la apariencia de un ser humano y actúe o sienta como él. Lo que la hace inteligente es que comprende e interactúa con su mundo valiéndose de un modelo de memoria jerárquico y es capaz de pensar sobre su mundo de modo análogo a como ustedes y yo reflexionamos sobre el nuestro. Como veremos, sus pensamientos y acciones podrían ser completamente diferentes de los de un ser humano, pero seguiría siendo inteligente. La inteligencia se mide por la capacidad predictiva de una memoria jerárquica, no por una conducta semejante a la humana.

\* \* \*

**P**restemos atención ahora al mayor desafío que arrostraremos cuando construyamos máquinas inteligentes: la creación de la memoria. Para lograrlo, necesitaremos componer grandes sistemas de memoria organizados jerárquicamente y que funcionen como la corteza cerebral. La capacidad y la conectividad supondrán grandes retos.

La *capacidad* es el primer asunto. Digamos que la corteza cerebral cuenta con treinta y dos billones de sinapsis. Si asumimos que cada una utiliza solo dos bits (que nos dan cuatro posibles valores por sinapsis) y cada byte tiene ocho bits (por lo tanto, un byte podría representar cuatro sinapsis), necesitaríamos unos ocho billones de bytes de memoria. Un disco duro de un ordenador personal actual tiene cien mil millones de bytes, así que necesitaríamos unos ochenta discos duros actuales para contar con la misma cantidad de memoria que la corteza cerebral. (No se preocupen por las cifras exactas, pues no son más que cálculos aproximados.) Lo importante es que esta cantidad de memoria es construible en el laboratorio. Por mucho que erremos en los cálculos, no es el tipo de máquina que cabe en un bolsillo o se puede incorporar a un tostador. De todos modos, lo

importante es que la cantidad de memoria requerida no se encuentra fuera de nuestro alcance, mientras que hace solo diez años sí lo habría estado. Nos ayuda el hecho de que no tenemos que recrear una corteza cerebral humana completa, pues mucho menos bastará para diversas aplicaciones.

Nuestras máquinas inteligentes necesitarán mucha memoria. Es probable que comencemos a construirlas utilizando discos duros o discos ópticos, pero acabaremos fabricándolas también de silicio. Los chips de silicio son pequeños, gastan poca energía y son resistentes. Y es solo cuestión de tiempo que se puedan fabricar chips de memoria de silicio con la capacidad precisa para construir máquinas inteligentes. De hecho, la memoria inteligente tiene una ventaja sobre la memoria informática convencional. La economía de la industria de los semiconductores se basa en el porcentaje de chips con errores. El porcentaje de chips buenos recibe el nombre de cosecha y determina si el diseño de uno particular puede fabricarse y venderse con beneficio. Como la posibilidad de error aumenta a medida que lo hace el chip, la mayoría de los actuales no exceden el tamaño de un sello postal. La industria no ha potenciado la cantidad de memoria en un único chip haciéndolo más grande, sino, en líneas generales, reduciendo sus rasgos particulares.

Pero los chips inteligentes de memoria tienen que tolerar los fallos. Recordemos que ningún componente único de nuestros cerebros contiene un elemento indispensable de datos. Nuestros cerebros pierden miles de neuronas cada día, pero nuestra capacidad mental va decreciendo a ritmo lento a lo largo de la vida adulta. Los chips de memoria inteligentes funcionarán según los mismos principios que la corteza cerebral, así que aunque un porcentaje de los elementos de la memoria resulten defectuosos, el chip seguirá siendo útil y rentable comercialmente. Es muy probable que la tolerancia a los errores de la memoria semejante al cerebro permita a los diseñadores construir chips que sean mucho más grandes y densos que los chips de memoria de los ordenadores actuales. El resultado es que tal vez contemos con un cerebro fabricado en silicio antes de lo que las tendencias presentes indican.

El segundo problema que debemos superar es la *conectividad*. Los cerebros reales poseen grandes cantidades de materia blanca subcortical. Como ya hemos señalado, la materia blanca está compuesta por millones de axones que fluyen de un lado a otro bajo la delgada lámina cortical y conectan entre sí las diferentes regiones de la jerarquía cortical. Una célula particular de la corteza cerebral puede conectarse con otras cinco o diez mil células. Este tipo de conexión paralela masiva es difícil o imposible de lograr utilizando las técnicas de fabricación de silicio tradicionales. Los chips de silicio se fabrican depositando unas cuantas capas de metal, separada cada una por una capa de aislamiento. (Este proceso no tiene nada que ver con las capas de la corteza cerebral.) Las capas de metal contienen los "cables" del chip, y como no pueden atravesar la que les corresponde, el

número total de conexiones es limitado. No va a servir para los sistemas de memoria semejantes al cerebro, en los que se necesitan millones de conexiones. Los chips de silicio y la materia blanca no son muy compatibles.

Se precisará mucha ingeniería y experimentación para resolver este problema, pero sabemos lo básico sobre el modo de solucionarlo. Los cables eléctricos envían señales mucho más deprisa que los axones de las neuronas. Un único cable de un chip puede compartirse y, por lo tanto, ser utilizado para muchas conexiones diferentes, mientras que en el cerebro cada axón pertenece a una sola neurona.

Un ejemplo del mundo real es el sistema telefónico. Si tendiéramos una línea de cada teléfono a otro, la superficie del globo quedaría enterrada bajo una selva de alambre de cobre. Lo que hacemos en su lugar es que todos los teléfonos compartan un número relativamente pequeño de líneas de alta capacidad. Este método funciona siempre que la capacidad de cada línea sea mucho mayor que la requerida para transmitir una sola conversación. El sistema telefónico cumple este requisito: un único cable de fibra óptica puede transportar a la vez un millón de conversaciones.

Los cerebros reales poseen axones exclusivos entre todas las células que hablan entre sí, pero podemos construir máquinas inteligentes que se asemejen más al sistema telefónico y compartan conexiones. Lo crean o no, algunos científicos han venido pensando cómo resolver el problema de conectividad del chip cerebral durante muchos años. Aunque el funcionamiento de la corteza cerebral continuaba siendo un misterio, los investigadores sabían que algún día desentrañaríamos el rompecabezas y entonces tendríamos que abordar el tema de la conectividad. No es preciso que analicemos aquí los diferentes planteamientos. Basta con decir que la conectividad podría ser el mayor obstáculo técnico que nos encontramos para construir máquinas inteligentes, pero tenemos que ser capaces de solventarlo.

Una vez que se han superado los retos tecnológicos, no hay problemas fundamentales que nos impidan construir sistemas verdaderamente inteligentes. Es cierto que existen multitud de temas que habrá que abordar para que estos sistemas sean pequeños, de bajo coste y gasten poca energía, mas nada se interpone en nuestro camino. Se necesitaron cincuenta años para pasar de los ordenadores del tamaño de una habitación a los que caben en el bolsillo, pero como partimos de una posición tecnológica avanzada, la transición a las máquinas inteligentes tiene que ser mucho más rápida.

### **¿Debemos Construir Máquinas Inteligentes?**

En el siglo XXI las máquinas inteligentes saldrán del reino de la ciencia ficción para convertirse en realidad. Antes de que llegue ese momento, debemos reflexionar sobre asuntos éticos, si los peligros posibles superarán a los beneficios probables.

La perspectiva de contar con máquinas capaces de pensar y actuar por su cuenta ha preocupado a la gente desde hace mucho tiempo, y es comprensible. Los nuevos campos del conocimiento y las tecnologías novedosas siempre han amedrentado cuando han aparecido. La creatividad humana nos hace imaginar formas terribles en las que una nueva tecnología puede apoderarse de nuestros cuerpos, volvernos inútiles o suprimir el valor mismo de la vida humana. Pero la historia muestra que esas oscuras imaginaciones casi nunca se cumplen del modo que esperamos. Cuando se produjo la Revolución Industrial, sentimos miedo de la electricidad (¿recuerdan a Frankenstein?) y los motores de vapor. La maquinaria que contaba con energía propia, que podía moverse de formas complejas, parecía milagrosa y a la vez potencialmente siniestra. Pero la electricidad y los motores de combustión interna ya no nos resultan extraños ni siniestros. Forman parte de nuestro entorno, igual que el aire y el agua.

Cuando comenzó la revolución de la información, enseguida empezamos a temer a los ordenadores. Había incontables relatos de ciencia ficción sobre ordenadores potentes o redes informáticas que de forma espontánea cobraban conciencia de sí mismos y luego se hacían los amos. Pero ahora que los ordenadores se han integrado en la vida diaria, ese temor parece absurdo. El ordenador de su casa o Internet tienen las mismas posibilidades de cobrar conciencia que una caja registradora.

Toda tecnología puede aplicarse para fines buenos o malos, por supuesto, pero algunas son más proclives al mal uso o la catástrofe que otras. La energía atómica es peligrosa tanto en la forma de cabezas nucleares como de centrales nucleares, porque un solo accidente o un solo mal uso podría dañar o matar a millones de personas. Y aunque la energía nuclear es valiosa, se dispone de alternativas. La tecnología vehicular puede adoptar la forma de tanques y reactores de combate, o la de coches y aviones de pasajeros, y un percance o mal uso puede causar daño a mucha gente. Pero cabría afirmar que los vehículos son más esenciales para la vida moderna y menos peligrosos que la energía nuclear. El daño causado por un solo mal uso de un avión es mucho menor que el de una bomba nuclear. Existen muchas tecnologías que son casi completamente beneficiosas. Los teléfonos son un ejemplo. Su tendencia a comunicar y unir a la gente sobrepasa con creces cualquier efecto negativo. Lo mismo es aplicable a la electricidad y a la ciencia de la salud pública. En mi opinión, las máquinas inteligentes van a ser una de las tecnologías menos peligrosas y más beneficiosas que hayamos desarrollado jamás.

Sin embargo, algunos pensadores, como Bill Joy, cofundador de Sun Microsystems, temen que desarrollemos nanorrobots inteligentes que se escapen de nuestro control, invadan la Tierra y la rehagan, átomo por átomo, según sus propios intereses. La imagen me trae a la mente esas escobas animadas por magia de *El aprendiz de brujo*, que se regeneran de sus astillas y trabajan incansables para provocar el desastre. En líneas similares, algunos optimistas de la inteligencia



artificial ofrecen profecías de extensión de la vida que resultan inquietantes. Por ejemplo, Ray Kurzweil habla del día en que los nanorrobots se nos introducirán en el cerebro para grabar cada sinapsis y cada comunicación y luego pasar toda la información a un superordenador, que se reconfigurará en nosotros. Nos convertiremos en una versión de *software* de nosotros mismos que será prácticamente inmortal. Estos dos temores sobre la inteligencia de las máquinas, el escenario de las máquinas inteligentes que hacen estragos y el escenario de la descarga de nuestros cerebros en un ordenador, parecen salir a la superficie una y otra vez.

Construir máquinas inteligentes no es lo mismo que construir máquinas autorreproductoras. No hay ninguna conexión lógica entre ambas cosas. Ni los cerebros ni los ordenadores pueden duplicarse de forma directa, y los sistemas de memoria semejantes a los cerebros no serán diferentes. Aunque uno de los puntos fuertes de las máquinas inteligentes será nuestra capacidad para producirlas en serie, es algo muy distinto a la duplicación al modo de las bacterias y los virus. La duplicación no requiere inteligencia, y la inteligencia no requiere duplicación.

Asimismo, dudo seriamente de que logremos copiar nuestras mentes en máquinas. En la actualidad, hasta donde sé, no existen métodos reales o imaginados capaces de grabar los miles de billones de detalles que nos crean a cada uno. Necesitaríamos grabar y recrear todo nuestro sistema nervioso y nuestro cuerpo, no solo la corteza cerebral. Y sería preciso comprender cómo funciona todo ello. Sin duda, algún día tal vez seamos capaces de conseguirlo, pero los retos se extienden mucho más allá de la comprensión del funcionamiento de la corteza cerebral. Imaginarse su algoritmo e incorporarlo a las máquinas desde cero es una cosa, pero escanear los tropecientos detalles operativos de un cerebro vivo y duplicarlos en una máquina es algo completamente diferente.

\* \* \*

**Más allá de la duplicación y el copiado de mentes, a la gente le preocupa otra cosa de las máquinas inteligentes. ¿Podrían amenazar de algún modo a grandes cantidades de población, como hacen las bombas nucleares? ¿Podría llevar su presencia a que pequeños grupos o individuos malévolos se volvieran muy poderosos? ¿O podrían volverse malas y actuar contra nosotros, como las máquinas implacables de las películas *Terminator* y *Matrix*?**

La respuesta a estas preguntas es no. Como aparatos de información, los sistemas de memoria semejantes a los cerebros se encontrarán entre las tecnologías más útiles que hemos desarrollado hasta el momento. Pero, al igual que los coches y los ordenadores, solo serán herramientas. Que vayan a ser inteligentes no significa que tendrán una habilidad especial para destruir la propiedad o manipular a la gente. Y del mismo modo que no pondríamos el control del arsenal nuclear del mundo bajo la autoridad de una persona o un ordenador, habremos de tener

cuidado para no depender demasiado de ellas, pues fallarán como lo hace toda la tecnología.

Esto nos lleva a la cuestión de la malevolencia. Algunas personas dan por sentado que ser inteligente equivale básicamente a tener mentalidad humana. Temen que las máquinas inteligentes se resientan de estar “esclavizadas”, porque los humanos aborrecen esa condición. Temen que las máquinas inteligentes intenten apoderarse del mundo, porque la gente inteligente a lo largo de la historia ha tratado de dominarlo. Pero estos temores descansan en una analogía falsa. Se basan en una refundición de la inteligencia —el algoritmo cortical— con los impulsos emocionales del cerebro viejo —cosas como temor, paranoia y deseo—. Pero las máquinas inteligentes no tendrán estas facultades. No tendrán ambición personal. No desearán riqueza, reconocimiento social o gratificación sensual. No tendrán apetitos, adicciones o trastornos de carácter. Las máquinas inteligentes no poseerán nada que se parezca a la emoción humana, a menos que nos tomemos el trabajo de diseñarlas a tal efecto. Sus aplicaciones más valiosas serán allí donde el intelecto humano encuentra dificultad, áreas en las que nuestros sentidos resultan inadecuados, o actividades que nos resultan aburridas. En general, estas actividades poseen poco contenido emocional.

Las máquinas inteligentes abarcarán desde sistemas sencillos de una sola aplicación hasta sistemas inteligentes superhumanos muy potentes; pero a menos que nos salgamos del camino para hacerlas semejantes a los humanos, no lo serán. Tal vez algún día tengamos que poner restricciones a lo que se puede hacer con ellas, mas es algo que está lejísimo y, cuando llegue, es probable que los asuntos éticos resulten relativamente sencillos comparados con cuestiones morales actuales como las que rodean a la genética y la tecnología nuclear.

### **¿Por qué Construir Máquinas Inteligentes?**

Analicemos ahora qué harán las máquinas inteligentes.

Me piden a menudo que dé charlas sobre el futuro de la informática móvil. El organizador de una conferencia me solicitará que describa cómo serán los ordenadores de bolsillo o los teléfonos móviles en cinco o veinte años. Quieren escuchar mi visión del futuro, pero no puedo darla. Para establecer mi premisa, una vez salí a escena con un sombrero de mago y una bola de cristal. Expliqué que nadie es capaz de ver el futuro con detalle. Las bolas de cristal son ficción, y cualquiera que pretenda saber con exactitud lo que pasará en los años venideros fallará con total seguridad. Lo más que cabe hacer es comprender tendencias generales. Si entendemos una idea general, seremos capaces de seguirla dondequiera que vaya mientras se despliegan los detalles.

El más famoso ejemplo de la tendencia de una tecnología es la ley de Moore. Gordon Moore acertó al predecir que el número de elementos de circuito que

podrían colocarse en un chip de silicio se duplicaría cada año y medio. No dijo si serían chips de memoria, unidades de procesamiento central u otra cosa. No indicó para qué tipo de productos se utilizarían los chips. No predijo si los chips se alojarían en plástico o cerámica, o si se pegarían a placas de circuito. No dijo nada sobre los diversos procesos empleados para fabricar los chips. Se limitó a expresar la tendencia más amplia que pudo, y acertó.

Hoy por hoy no podemos predecir los usos finales de las máquinas inteligentes, pues no hay modo de acertar con los detalles. Si yo u otra persona predecimos minuciosamente lo que harán esas máquinas, será inevitable que nos equivoquemos. No obstante, se puede hacer algo más que limitarnos a encogernos de hombros. Existen dos líneas de pensamiento útiles. Una es conjeturar los usos a muy corto plazo de los sistemas de memoria semejantes al cerebro: las cosas obvias pero menos interesantes que se intentarán primero. El segundo planteamiento es reflexionar sobre las tendencias a largo plazo, como la ley de Moore, que pueden ayudarnos a imaginar las aplicaciones que probablemente formarán parte de nuestro futuro.

Comencemos con algunas aplicaciones a corto plazo. Son las cosas que parecen obvias, como reemplazar los tubos de una radio por transistores o construir calculadoras con un microprocesador. Y podemos empezar observando algunas áreas que la inteligencia artificial intentó desentrañar pero no pudo: reconocimiento de voz, visión y coches inteligentes.

\* \* \*

**S**i alguna vez ha intentado usar *software* de reconocimiento de voz para introducir texto en un ordenador personal, sabe lo absurdo que puede resultar. Al igual que la habitación china de Searle, el ordenador no tiene conocimiento de lo que se le dice. Las pocas veces que he probado esos productos acabé frustrado. Si había algún ruido en la habitación, desde un lápiz que se cae hasta alguien que me hablara, aparecían palabras adicionales en mi pantalla. Los índices de error eran elevados. Con frecuencia las palabras que el *software* pensaba que yo decía carecían de sentido. "Recuerda pedirle a Mary que el tarro está preparado para que lo paseen." Un niño se daría cuenta de que era un error, pero no el ordenador. De forma similar, los denominados interfaces de lenguaje natural han sido una meta para los científicos informáticos durante años. La idea es que seamos capaces de decirle a un ordenador u otro electrodoméstico lo que queremos en lenguaje llano y dejar que la máquina realice el trabajo. A un asistente digital personal o PDA se le podría indicar: "Traslada el partido de baloncesto de mi hija al domingo a las diez de la mañana". Este tipo de cosas ha sido imposible hacerlas bien con la inteligencia artificial tradicional. Aun cuando el ordenador pudiera reconocer cada palabra, para completar la tarea necesitaría saber dónde va al colegio tu hija, que probablemente quieres decir el próximo domingo y tal vez qué es un partido de baloncesto, porque la cita podría decir solo

“Menlo contra St. Joe”. O quizá queramos que un ordenador escuche una emisión de radio en busca de la mención de un producto particular, mas el anunciador radiofónico describe el producto sin utilizar su nombre. Usted y yo sabríamos de qué está hablando, pero no un ordenador.

Estas y otras aplicaciones requieren que la máquina sea capaz de escuchar el lenguaje hablado. Pero los ordenadores no logran realizar estas tareas porque no entienden lo que se dice. Casan patrones auditivos con plantillas de palabras por el tono, sin saber qué significan las palabras. Imaginemos que aprendiéramos a reconocer los sonidos de las palabras particulares de una lengua extranjera, pero no su significado, y nos pidieran que transcribiéramos una conversación en dicha lengua. La conversación va discurriendo y no tenemos idea de qué trata, mas intentamos recoger las palabras aisladas. Sin embargo, las palabras se solapan e interfieren, y se pierden partes del sonido por el ruido. Nos resultaría extremadamente difícil separar palabras y reconocerlas. El *software* de reconocimiento de voz lucha contra estos obstáculos en la actualidad. Los ingenieros han descubierto que utilizando probabilidades de transiciones de palabras pueden mejorar algo su precisión. Por ejemplo, emplean reglas gramaticales para decidir entre dos homónimos. Es una forma muy simple de predicción, pero los sistemas siguen resultando absurdos. Solo logran éxito en situaciones muy reducidas en las que el número de palabras que se podrían pronunciar en un momento dado sea limitado. Sin embargo, los seres humanos realizan con facilidad muchas tareas relacionadas con el lenguaje porque nuestra corteza cerebral entiende no solo las palabras, sino las oraciones y el contexto en el que se expresan. Adelantamos ideas, expresiones y palabras particulares. Nuestro modelo cortical del mundo lo hace de forma automática.

Así pues, cabe esperar que los sistemas de memoria semejantes a la corteza cerebral transformen el reconocimiento de voz falible en una comprensión ajustada del habla. En lugar de programar probabilidades para transiciones de una sola palabra, la memoria jerárquica rastreará acentos, palabras, expresiones e ideas, y los utilizará para interpretar lo que se está diciendo. Al igual que una persona, esa máquina inteligente podría distinguir entre varios hechos del habla; por ejemplo, una discusión entre usted y un amigo en la habitación, una conversación telefónica y las órdenes de edición para un libro. No será fácil construir dichas máquinas. Para entender de lleno el lenguaje humano, la máquina tendrá que experimentar y aprender lo que los humanos hacen. Por lo tanto, aunque quizá tardemos algunos años en conseguir construir una máquina inteligente que entienda el lenguaje tan bien como ustedes y yo, a corto plazo seremos capaces de mejorar los resultados de los sistemas de reconocimiento de voz existentes empleando memorias semejantes a la corteza cerebral.

La visión ofrece otro conjunto de aplicaciones que la inteligencia artificial ha sido incapaz de lograr pero que los sistemas inteligentes reales deben manejar. En la

actualidad no hay una máquina que pueda mirar una escena natural —el mundo ante nuestros ojos— o una foto usando una cámara y describir lo que ve. Existen unas cuantas aplicaciones de visión en máquinas que funcionan en campos muy restringidos, como la alineación visual de chips en una placa de circuito o la correlación de rasgos faciales con bases de datos, mas es imposible para un ordenador identificar una variedad de objetos o analizar una escena de modo más general. No nos cuesta nada mirar alrededor de una habitación y encontrar un lugar donde sentarnos, pero no le pidamos a un ordenador que lo haga. Imaginemos que miramos la pantalla de vídeo de una cámara de seguridad. ¿Podríamos notar la diferencia entre alguien que llama a la puerta sosteniendo un regalo y alguien que golpea la puerta con una palanca? Por supuesto que sí, mas la distinción está mucho más allá de la capacidad del *software* actual. En consecuencia, contratamos a personas para que echen un ojo a las pantallas de las cámaras de seguridad las veinticuatro horas en busca de algo sospechoso. A los vigilantes humanos les resulta difícil permanecer alerta, mientras que una máquina inteligente podría realizar la tarea sin cansarse. Muchas situaciones que dependen de la agudeza visual requieren además la comprensión de una escena compleja. Las máquinas inteligentes son el único modo de atenderlas.

Por último, analicemos el transporte. Los coches se están volviendo muy complicados. Tienen GPS para trazar la ruta de A a B, sensores para encender las luces cuando anochece, acelerómetros para desplegar los *airbags* y sensores de proximidad para indicarnos que estamos a punto de darnos con algo. Incluso hay coches que pueden conducirse de forma autónoma en autopistas especiales o en condiciones ideales, aunque no se comercializan. Pero para conducir un coche con seguridad y eficacia en todo tipo de carreteras y condiciones de tráfico se requiere más que unos cuantos sensores y circuitos de control de realimentación. Para ser un buen conductor se debe comprender el tráfico, a los restantes conductores, el funcionamiento de los coches, los semáforos y otras múltiples cosas. Es preciso ser capaz de entender las señales de peligro o darse cuenta de que otro coche es conducido de modo peligroso. Es necesario ver el intermitente de otro coche y prever que es probable que cambie de carril, o si la luz permanece encendida varios minutos, valorar que es probable que el conductor no se haya dado cuenta y, por lo tanto, no cambiará de carril. Es preciso reconocer que una columna de humo en la lejanía podría significar que ha ocurrido un accidente y, por consiguiente, se debe reducir la velocidad. Un conductor que ve un balón cruzar la calle, y piensa de forma automática que un niño correrá a cogerlo, se detiene por intuición.

Digamos que queremos construir un coche verdaderamente inteligente. Lo primero que haríamos es seleccionar un conjunto de sensores que le permitieran sentir su mundo. Podríamos comenzar con una cámara para ver, quizá cámaras múltiples delante y detrás, y micrófonos para oír, pero tal vez también querríamos ponerle radar o sensores de ultrasonido capaces de determinar con precisión el alcance y

la velocidad de otros objetos en condiciones de escasa visibilidad. Lo que quiero señalar es que no tenemos que recurrir a los sentidos que usan los seres humanos, ni restringirnos a ellos. El algoritmo cortical es flexible y, siempre que diseñemos nuestro sistema de memoria jerárquico de forma adecuada, debe funcionar prescindiendo del tipo de sensores que instalemos. En teoría, nuestro coche podría aventajarnos en la percepción del mundo del tráfico porque su conjunto de sentidos se ha elegido para cumplir esa tarea. Los sensores se acoplarían a un sistema de memoria jerárquico con una amplitud suficiente. Los diseñadores del coche entrenarían su memoria exponiéndola a las condiciones del mundo real para que aprenda a construir un modelo de su mundo del mismo modo que lo hacen los humanos, solo que en un campo más limitado. (Por ejemplo, el coche necesita saber de carreteras, pero no de ascensores y aviones.) Su memoria aprendería la estructura jerárquica del tráfico y las carreteras para que entienda y prevea lo que es probable que suceda en su mundo de automóviles en movimiento, señales de tráfico, obstáculos e intersecciones. Los ingenieros podrían diseñar el sistema de memoria para que conduzca realmente el coche o se limite a explorar lo que sucede cuando conducimos nosotros. Podría dar consejos o asumir el control en situaciones extremas, como un conductor del asiento trasero al que no le tomamos a mal que lo haga. Una vez que la memoria está bien entrenada y el coche es capaz de comprender y resolver lo que pueda pasar, los ingenieros tendrían la opción de fijarla para que todos los coches que salgan de la cadena de montaje se comporten del mismo modo, o podrían diseñarla para que continuara aprendiendo una vez que se ha vendido el coche. Y como en el caso de un ordenador, pero no de un ser humano, la memoria podría reprogramarse con una versión actualizada si las nuevas condiciones lo requieren.

No digo que vayamos a construir coches inteligentes o máquinas que comprendan el lenguaje y la visión, mas son buenos ejemplos del tipo de aparatos que podríamos investigar y desarrollar, y que parece posible construir.

\* \* \*

**P**or lo que a mí se refiere, siento poco interés por las aplicaciones obvias de las máquinas inteligentes. Creo que el beneficio y estímulo verdaderos de una nueva tecnología es encontrarle usos que antes resultarían inconcebibles. ¿De qué formas nos sorprenderán las máquinas inteligentes y qué capacidades fantásticas surgirán con el tiempo? Estoy seguro de que las memorias jerárquicas, al igual que el transistor y el microprocesador, transformarán nuestras vidas a mejor de modos increíbles; pero ¿cómo? Una manera de atisbar el futuro de las máquinas inteligentes es pensar en aspectos de la tecnología que sean escalables; es decir, qué atributos de las máquinas inteligentes se irán volviendo cada vez más baratos, cada vez más rápidos o cada vez más pequeños. Las cosas que crecen a tasas exponenciales sobrepasan enseguida nuestra imaginación y es muy probable que

desempeñen un papel clave en las evoluciones más radicales de la tecnología futura.

Entre los ejemplos de tecnologías que han mejorado de forma exponencial durante muchos años se incluye el chip de memoria de silicio, el disco duro, las técnicas de secuenciación del ADN y la transmisión por fibra óptica. Estas tecnologías han sido la base de muchos productos y empresas nuevos. De un modo diferente, el *software* también es escalable. Un programa deseado, una vez que se ha escrito, puede copiarse infinitamente casi sin coste alguno.

En contraste, algunas tecnologías, como las baterías, los motores y la robótica tradicional, no son escalables. A pesar de multitud de esfuerzos y mejoras constantes, un brazo robótico construido hoy no es mucho mejor que otro construido hace diez años. Los avances de la robótica son graduales y modestos, y se hallan muy lejos de las curvas de crecimiento exponencial del diseño de chips o la proliferación de *software*. Un brazo de robot construido en 1985 por un millón de dólares no puede construirse hoy mil veces más fuerte por solo diez dólares. Del mismo modo, las baterías de hoy no son muchísimo mejores que las de hace diez años. Cabría afirmar que son dos o tres veces mejores, pero no mil o diez mil veces mejores, y el progreso avanza muy poco a poco. Si la capacidad de las baterías aumentara al mismo ritmo que la de los discos duros, los teléfonos móviles y otros aparatos electrónicos, jamás tendrían que recargarse y los coches eléctricos ligeros que recorrieran 1.600 kilómetros por carga se volverían comunes.

Así pues, nos corresponde pensar qué aspectos de los sistemas de memoria semejante al cerebro acabarán sobrepasando de forma espectacular a nuestros cerebros biológicos. Estos atributos nos sugerirán dónde terminará la tecnología. Veo cuatro atributos que superarán nuestras facultades: velocidad, capacidad, posibilidad de duplicación y sistemas sensoriales.

### ***Velocidad***

Mientras que las neuronas funcionan en el orden de milisegundos, el silicio lo hace en el de nanosegundos (y sigue cobrando velocidad), lo que supone una diferencia de un millón de veces o seis órdenes de magnitud. La diferencia de velocidad entre las mentes orgánicas y las basadas en silicio será de gran importancia. Las máquinas inteligentes serán capaces de pensar un millón de veces más deprisa que el cerebro humano. Dicha mente podría leer bibliotecas enteras o estudiar enormes compilaciones de datos complicados en solo minutos, extrayendo la misma comprensión. No hay nada de magia en ello. Los cerebros biológicos evolucionaron con dos limitaciones relacionadas con el tiempo. Una es la velocidad a la que las células pueden hacer las cosas y la otra es la velocidad a la que cambia el mundo. Tal vez no sea demasiado útil para un cerebro biológico pensar un millón de veces más deprisa si el mundo que lo rodea es lento de por sí. Pero no hay nada en el algoritmo cortical que diga que debe operar siempre despacio.

Si una máquina inteligente conversara o interactuara con un humano, tendría que aminorar la marcha para funcionar a velocidad humana. Si lee un libro pasando páginas, habría un límite a la velocidad con que podría hacerlo. Pero cuando se conecta con el mundo electrónico, podría funcionar mucho más deprisa. Dos máquinas inteligentes podrían mantener una conversación un millón de veces más deprisa que dos humanos. Imaginemos el avance de una máquina inteligente que resolviera problemas matemáticos o científicos un millón de veces más deprisa que un ser humano. En diez segundos podría dedicar más reflexión a un problema de lo que nosotros lograríamos en un mes. Sin duda, mentes que no se cansan ni aburren jamás con esa velocidad ingente resultarán útiles de formas que todavía no podemos imaginar.

### ***Capacidad***

Pese a la impresionante capacidad de memoria de una corteza cerebral humana, las máquinas inteligentes que construyamos pueden sobrepasarla con creces. El tamaño de nuestro cerebro se ha visto limitado por varios factores biológicos, entre los que se incluyen la proporción del cráneo del niño en relación con el diámetro pélvico de la madre, el alto coste metabólico del funcionamiento cerebral (nuestro cerebro supone en torno al 2 por 100 del peso corporal, pero usa casi el 20 por 100 del oxígeno que respiramos), y la lentitud de las neuronas. Mas podemos construir sistemas de memoria inteligentes de cualquier tamaño, y aportar previsión y una intención específica a los detalles del diseño. La capacidad de la corteza cerebral humana tal vez resulte relativamente modesta dentro de unas décadas.

Al construir máquinas inteligentes, podríamos aumentar su capacidad de memoria de varios modos. Si se añade profundidad a la jerarquía, se conseguirá una comprensión más penetrante, pues se verán patrones de órdenes superiores. Si se amplía la capacidad de las regiones, la máquina recordará más detalles, o percibirá con mayor agudeza, del mismo modo que una persona ciega tiene un sentido del tacto o el oído más finos. Y si extendemos las jerarquías sensoriales, construirá mejores modelos del mundo, como expondré en breve.

Será interesante comprobar si existe un límite máximo para el tamaño que puede alcanzar un sistema de memoria inteligente y en qué dimensiones. Parece razonable pensar que el aparato se recargaría demasiado para resultar útil o incluso fallaría cuando se acerque a cierto límite teórico. Quizá el cerebro humano ya se encuentra cerca del tamaño máximo teórico, pero no me parece probable. Los cerebros humanos se agrandaron en fecha muy reciente del período evolutivo y no hay nada que sugiera que nos hallemos en un tamaño máximo estable. Prescindiendo de cuál resulte ser la capacidad máxima de un sistema de memoria inteligente, es casi seguro que el cerebro humano no la ha alcanzado. Es probable que ni siquiera se aproxime.



Un modo de considerar lo que podrían hacer estos sistemas es observar los límites del rendimiento humano conocido. Sin duda, Einstein era inteligentísimo, pero su cerebro seguía siendo un cerebro. Cabe suponer que su extraordinaria inteligencia fue en buena medida producto de las diferencias físicas entre su cerebro y el cerebro humano típico. Lo que hizo a Einstein tan raro fue que nuestros genes no suelen producir cerebros como el suyo. Sin embargo, cuando diseñamos cerebros en silicio, podemos construirlos como queramos. Podrían ser capaces de alcanzar el elevado nivel intelectual de Einstein, o incluso sobrepasarlo. En el otro extremo, las personas que padecen el síndrome del sabio (*savant syndrome*) pueden arrojar luz sobre otras posibles dimensiones de la inteligencia. Los individuos con dicho síndrome son retrasados mentales que muestran habilidades notables, como memorias casi fotográficas o capacidad para realizar cálculos matemáticos difíciles a gran velocidad. Sus cerebros, aunque no son típicos, siguen siendo cerebros y funcionan con el algoritmo cortical. Si un cerebro atípico puede poseer habilidades de memoria asombrosas, en teoría podríamos añadir dichas habilidades a nuestros cerebros artificiales. Estos extremos de la capacidad mental humana no solo indican lo que debería ser posible recrear, sino que representan la dirección en la que es probable que superemos los mejores resultados humanos.

### ***Posibilidad de Duplicación***

Cada nuevo cerebro orgánico debe crecer y entrenarse desde cero, proceso que supone décadas en los seres humanos. Cada humano ha de descubrir lo esencial de la coordinación de los miembros y grupos musculares del cuerpo, del equilibrio y del movimiento, y aprender las propiedades generales de multitud de objetos, animales y otras personas; los nombres de las cosas y la estructura del lenguaje; y las reglas de la familia y la sociedad. Una vez que se dominan esos elementos básicos, comienzan años y años de enseñanza formal. Todas las personas deben recorrer el mismo conjunto de curvas de aprendizaje en la vida —aunque ya las hayan transitado muchos otros innumerables veces— para construir un modelo del mundo en la corteza cerebral.

Las máquinas inteligentes no necesitan pasar por esta larga curva de aprendizaje, puesto que los chips y el resto de componentes pueden duplicarse de forma infinita y los contenidos transferirse con facilidad. En este sentido, las máquinas inteligentes se duplicarían como el *software*. Una vez que se ha puesto a punto y entrenado un sistema prototipo de forma satisfactoria, podría copiarse tantas veces como se quiera. Puede que lleve años de diseño de chips, configuración de *hardware*, entrenamiento y tanteo perfeccionar el sistema de memoria de un coche inteligente, pero una vez que se ha logrado el producto final, se producirá en serie. Como ya he mencionado, podríamos decidir que las copias continuaran aprendiendo o no. Para algunas aplicaciones preferiremos que nuestras máquinas inteligentes se limiten a operar de una forma probada y conocida. Una vez que un coche inteligente sabe todo lo que necesita, no querríamos que desarrollara malos

hábitos o llegara a creer en alguna falsa analogía que le parezca ver. Además, esperaríamos que todos los coches de factura similar se comporten de forma parecida. Pero para otras aplicaciones preferiremos que nuestros sistemas de memoria semejantes al cerebro sean plenamente capaces de seguir aprendiendo. Por ejemplo, una máquina inteligente diseñada para descubrir pruebas matemáticas necesitará la capacidad de aprender de la experiencia para aplicar percepciones antiguas a nuevos problemas, y en general para ser flexible y de duración indefinida.

Será posible compartir componentes de aprendizaje del mismo modo que compartimos componentes de *software*. Una máquina inteligente de un diseño particular podría reprogramarse con un nuevo juego de conexiones que lleven a una conducta diferente, como si se descargara un nuevo conjunto de conexiones en su cerebro para que pasara de forma inmediata de anglohablante a francohablante, o de profesor de ciencias políticas a musicólogo. Se podría intercambiar y aprovechar el trabajo de los demás. Pongamos que he desarrollado y entrenado una máquina con un sistema de visión superior y otra persona ha desarrollado y entrenado una máquina con un sentido del oído superior. Con el diseño adecuado, podríamos combinar lo mejor de ambos sistemas sin tener que efectuar un nuevo entrenamiento de arriba abajo. Compartir la pericia de este modo es imposible para los humanos. La construcción de máquinas inteligentes podría evolucionar siguiendo las mismas líneas que la industria informática, con grupos de personas dedicadas a entrenar a las máquinas inteligentes para que tengan conocimientos y habilidades especializadas, y vendiendo e intercambiando las configuraciones de memoria resultantes. La reprogramación de una máquina inteligente no será demasiado diferente de la instalación de un nuevo videojuego o una pieza de *software*.

### ***Sistemas Sensoriales***

Los seres humanos cuentan con un puñado de sentidos que están profundamente arraigados en nuestros genes, nuestros cuerpos y el entramado subcortical de nuestros cerebros. No podemos cambiarlos. A veces empleamos la tecnología para aumentarlos, como es el caso de las gafas de visión nocturna, el radar y el telescopio espacial *Hubble*. Estos instrumentos de alta tecnología son ingeniosos dispositivos de interpretación de datos, no nuevos modos de percepción. Convierten información que no podemos ver en indicaciones visuales o auditivas que somos capaces de interpretar. De todos modos, es la asombrosa flexibilidad de nuestro cerebro la que hace posible que miremos la pantalla de un radar y comprendamos lo que representa. Muchas especies de animales poseen verdaderos sentidos diferentes, como la ecolocalización de los murciélagos y delfines, la capacidad de las abejas de ver la luz polarizada y ultravioleta, y el sentido de campo eléctrico de algunos peces.

Nuestras máquinas inteligentes podrían percibir el mundo a través de cualquier sentido encontrado en la Naturaleza, así como de nuevos sentidos de diseño puramente humano. El sonar, el radar y la visión infrarroja son ejemplos evidentes del tipo de sentidos no humanos que tal vez nos gustaría que poseyeran nuestras máquinas inteligentes. Pero solo son el comienzo.

Mucho más interesante es el modo como las máquinas inteligentes experimentan mundos de sensaciones ajenos y genuinamente exóticos. Como hemos visto, el algoritmo cortical se ocupa sobre todo de encontrar patrones en el mundo. No tiene preferencia por los orígenes físicos de esos patrones. Siempre que las entradas a la corteza cerebral no sean aleatorias y tengan una cierta riqueza o estructura estadística, un sistema inteligente formará memorias invariables y predicciones al respecto. No hay razón para que estos patrones de entrada sean análogos a los sentidos de los animales, o se deriven del mundo real. Sospecho que es en el campo de los sentidos exóticos donde se encuentran los usos revolucionarios de las máquinas inteligentes.

Por ejemplo, podríamos diseñar un sistema sensorial que abarque el globo. Imaginemos sensores climáticos espaciados cada unos cien kilómetros por un continente. Estos sensores serían análogos a las células de la retina. En un momento dado, dos sensores climáticos adyacentes tendrán una alta correlación en su actividad, al igual que dos células adyacentes de una retina. Hay grandes fenómenos climáticos, como las tormentas y los frentes, que se mueven y cambian con el tiempo, al igual que hay fenómenos visuales que también se comportan así. Al unir este despliegue sensorial a una gran memoria semejante a la corteza cerebral, capacitaríamos al sistema para aprender a predecir el clima, así como ustedes y yo aprendemos a reconocer fenómenos visuales y a predecir cómo evolucionarán a lo largo del tiempo. El sistema vería patrones climáticos locales, patrones climáticos mayores y patrones que existen a lo largo de décadas, años u horas. Al colocar sensores muy cerca en algunas regiones, crearíamos lo equivalente a una fovea, con lo que permitiríamos a nuestro cerebro climático inteligente comprender y predecir microclimas. Nuestro cerebro climático entendería y pensaría sobre sistemas climáticos globales como ustedes y yo pensamos y entendemos objetos y personas. Los meteorólogos pretenden hacer algo similar en la actualidad. Reúnen registros de diversos lugares y emplean superordenadores para simular el clima y predecir el futuro. Pero este planteamiento —que es diferente en lo básico al modo como funcionaría una máquina inteligente— se asemeja a la manera de jugar al ajedrez de un ordenador —sin reflexión y sin entender—, mientras que nuestra máquina climática inteligente está más próxima a como juega al ajedrez un ser humano, reflexionando y comprendiendo. La máquina climática inteligente descubriría patrones que se les han escapado a los humanos. El fenómeno conocido como *El Niño* no se descubrió hasta la década de 1960. Nuestro cerebro climático podría encontrar más patrones como ese, o aprender a predecir tornados o monzones

mucho mejor que los humanos. Es difícil poner grandes cantidades de datos sobre el clima de una forma que los humanos puedan entender de inmediato; en cambio, nuestro cerebro climático sentiría y pensaría sobre el clima de forma directa.

Otros sistemas sensoriales distribuidos ampliamente nos permitirían construir máquinas inteligentes que comprendieran y predijeran las migraciones animales, los cambios demográficos y la propagación de la enfermedad. Imaginemos que contamos con sensores distribuidos en la red de energía eléctrica de un país. Una máquina inteligente adosada a estos sensores observaría el flujo y reflujo del consumo eléctrico, del mismo modo que ustedes y yo vemos el ajeteo del tráfico en una carretera, o el movimiento de la gente en un aeropuerto. Mediante la exposición repetida, los humanos aprenden a predecir estos patrones; no hay más que preguntarle a un empleado que va y viene al trabajo en coche o a un guarda de seguridad del aeropuerto. De igual modo, nuestro monitor inteligente de redes eléctricas sería capaz de predecir demandas de energía o situaciones peligrosas propensas a llevar a un apagón mejor que un humano. Podríamos combinar los sensores para el clima y la demografía humana con el fin de prever el descontento político, las hambrunas o el surgimiento de enfermedades. Al igual que un diplomático superdotado, las máquinas inteligentes podrían desempeñar un papel en la reducción del conflicto y el sufrimiento humanos. Tal vez piensen que las máquinas inteligentes necesitarían emociones para prever los patrones que supongan conducta humana, pero yo no lo creo. No hemos nacido con un juego de cultura, un juego de valores y un juego de religión; todo eso lo aprendemos. Y del mismo modo que yo puedo aprender a entender las motivaciones de personas con valores diferentes a los míos, las máquinas inteligentes pueden comprender las motivaciones y emociones humanas aunque ellas no las tengan.

Podríamos crear sentidos que se ocupen de entidades diminutas. Es en teoría posible contar con sensores que puedan representar patrones en células o moléculas grandes. Por ejemplo, un importante reto actual es comprender cómo es posible predecir la forma de la molécula de una proteína partiendo de la secuencia de aminoácidos que comprende dicha proteína. Ser capaces de predecir cómo se pliegan y actúan las proteínas llevaría al desarrollo de medicinas y curas para muchas enfermedades. Los ingenieros y científicos han creado modelos de proteínas visuales tridimensionales en un esfuerzo por predecir cómo se comportan estas moléculas complejas, pero la tarea ha resultado demasiado difícil. Sin embargo, una máquina superinteligente con un conjunto de sensores afinados de forma específica para esta misión tal vez hallarían la respuesta. Si parece exagerado, recordemos que no nos sorprendería que los humanos pudieran resolver el problema. Nuestra incapacidad para desentrañarlo tal vez tenga que ver sobre todo con la falta de adecuación entre los sentidos humanos y los fenómenos físicos que deseamos entender. Las máquinas inteligentes pueden tener sentidos creados por encargo y una memoria mayor que la humana, lo cual les permitiría resolver problemas imposibles para nosotros.

Con los sentidos adecuados y una leve reestructuración de la memoria cortical, nuestras máquinas inteligentes tal vez vivan y piensen en mundos virtuales usados en la matemática y la física. Por ejemplo, muchos empeños matemáticos y científicos requieren la comprensión del comportamiento de los objetos en mundos que tienen más de tres dimensiones. Los teóricos de las cuerdas, que estudian la misma naturaleza del espacio, piensan en un Universo con diez o más dimensiones. A los seres humanos les cuesta mucho reflexionar sobre problemas matemáticos en cuatro o más dimensiones. Tal vez una máquina inteligente con el diseño adecuado podría comprender espacios de dimensiones elevadas y, por lo tanto, ser apta para predecir cómo se comportan.

Por último, podríamos unir un grupo de sistemas inteligentes en una gran jerarquía, del mismo modo que nuestra corteza cerebral reúne el oído, el tacto y la visión ascendiendo en la jerarquía cortical. Dicho sistema aprendería de forma automática a modelar y predecir los patrones de pensamiento en poblaciones de máquinas inteligentes. Con medios de comunicación distribuidos como Internet, las máquinas inteligentes individuales podrían repartirse por todo el globo. Las jerarquías mayores aprenden patrones más profundos y ven analogías más complejas.

El objetivo de estas reflexiones es ilustrar que hay muchos aspectos en los que las máquinas semejantes al cerebro podrían sobrepasar nuestras capacidades de forma espectacular. Serían capaces de pensar y aprender un millón de veces más deprisa que nosotros, recordar ingentes cantidades de información detallada o ver patrones increíblemente abstractos. Pueden tener sentidos más sensibles que los nuestros, sentidos que estén distribuidos, o sentidos para fenómenos muy pequeños. Pueden pensar en tres, cuatro o más dimensiones. Ninguna de estas interesantes posibilidades depende de que las máquinas inteligentes imiten a los humanos o actúen como ellos, y no suponen una robótica compleja.

Ahora podemos ver plenamente cómo el test de Turing, al hacer equivalentes inteligencia y conducta humanas, limitó nuestra visión de lo que es posible. Entendiendo primero qué es la inteligencia, podemos construir máquinas inteligentes que son mucho más valiosas que la simple copia de la conducta humana. Nuestras máquinas inteligentes serán herramientas asombrosas y extenderán de forma espectacular nuestro conocimiento del Universo.

\* \* \*

¿Cuánto falta para que algo de esto se haga realidad? ¿Estaremos construyendo máquinas inteligentes en cincuenta, veinte o cinco años? Hay un refrán en el mundo de la alta tecnología que dice que el cambio tarda más de lo que se espera a corto plazo, pero ocurre más deprisa de lo que se espera a largo plazo. Lo he visto muchas veces. Alguien se levanta en una conferencia, anuncia una nueva tecnología y declara que estará en todos los hogares en cuatro años. Resulta que

el orador se equivoca. Cuatro años se convierten en ocho, y la gente comienza a pensar que nunca sucederá. Justo entonces, cuando parece que toda la idea era un callejón sin salida, empieza a despegar y se convierte en una gran sensación. Algo similar es probable que ocurra en el negocio de las máquinas inteligentes. El avance parecerá lento al principio, pero luego despegará con rapidez.

En las conferencias de neurociencia, me gusta pasearme por la sala y pedirle a cada participante que me exprese su opinión sobre cuánto tiempo pasará para que tengamos una teoría sobre la corteza cerebral que funcione. Unos cuantos — menos del 5 por 100— contestan “nunca” o “ya tenemos una” (respuestas sorprendentes, puesto que se ganan la vida con eso). Otro 5 por 100 afirma que cinco o diez años. La mitad restante dice de diez a quince años, o “en el transcurso de mi vida”. Los que quedan piensan que cincuenta o cien años o “no en el transcurso de mi vida”. Yo me coloco con los optimistas. Hemos estado viviendo en el período “lento” durante décadas, por eso a muchas personas les parece que el avance en la neurociencia teórica y las máquinas inteligentes se ha estancado. Si se juzga el avance realizado en los últimos treinta años, es natural suponer que no estamos nada cerca de una respuesta. Pero creo que nos hallamos en el momento decisivo y el asunto está a punto de despegar.

Es posible acelerar el futuro, acercar más el momento decisivo al presente. Una de las metas de este libro es convencerles de que, con el marco teórico correcto, podemos realizar un rápido progreso en el entendimiento de la corteza cerebral, que con el marco de memoria-predicción como guía podemos descifrar los detalles del funcionamiento y modo de pensar del cerebro. Este es el conocimiento que necesitamos para construir máquinas inteligentes. Si se trata del modelo acertado, el progreso puede avanzar con rapidez.

Así pues, aunque me cuesta predecir cuándo se hará realidad la era de las máquinas inteligentes, creo que si hoy se dedica la gente suficiente a resolver este problema, podremos ser capaces de crear prototipos y simulaciones corticales útiles en unos cuantos años. Antes de una década espero que las máquinas inteligentes sean una de las áreas más apasionantes de la tecnología y la ciencia. Me resisto a ser más específico, porque sé lo fácil que es subestimar el tiempo que se necesita para que algo importante suceda. Así pues, ¿por qué me muestro tan optimista acerca del rápido avance en la comprensión del cerebro y la construcción de máquinas inteligentes? Mi confianza proviene en buena parte del gran tiempo que ya he dedicado a trabajar sobre el problema de la inteligencia. Cuando me enamoré por primera vez de los cerebros en 1979, me pareció que resolver el rompecabezas de la inteligencia era algo que podría lograr en el transcurso de mi vida. Con el paso de los años, he observado con cuidado el declive de la inteligencia artificial, el ascenso y caída de las redes neuronales y la era del cerebro en la década de 1990. He visto cómo han evolucionado las actitudes hacia la biología teórica, y la neurociencia teórica en particular. He visto cómo han

llegado al lenguaje de la neurociencia las ideas de predicción, representación jerárquica y tiempo. He visto cómo avanzaba mi comprensión y la de mis colegas. Hace dieciocho años que me interesé por el papel de la predicción, y desde entonces en cierto modo he venido probándolo. Como he estado sumergido en los campos de la neurociencia y la informática durante más de dos décadas, tal vez mi cerebro ha construido un modelo de alto nivel sobre cómo ocurre el cambio tecnológico y científico, y ese modelo predice un progreso rápido. Ahora es el momento decisivo.

## Epílogo

**El** astrónomo Carl Sagan solía afirmar que la comprensión de algo no disminuye su prodigio y misterio. Mucha gente teme que la comprensión científica suponga un trueque con el prodigio, como si el saber le sorbiera a la vida su sabor y color. Pero Sagan estaba en lo cierto. La verdad es que con el entendimiento nos sentimos más cómodos con nuestro papel en el Universo, y a la vez este se vuelve todavía más colorido y misterioso. Ser una manchita diminuta en un Cosmos infinito, vivo, inteligente y creativo es mucho más maravilloso que vivir en una Tierra plana y limitada en el centro de un Universo pequeño. Comprender cómo funcionan nuestros cerebros no disminuye el prodigio y misterio del Universo, nuestras vidas o nuestro futuro. Nuestro asombro no hará más que aumentar cuando apliquemos este conocimiento a entendernos a nosotros mismos, construir máquinas inteligentes y entonces adquirir más conocimiento.

Al aceptar el reto, me acuerdo del físico Erwin Schrödinger, que en 1944 escribió un delgado volumen titulado *¿Qué es la vida?* en el que invitaba a los jóvenes científicos a tener presente que el funcionamiento de un organismo requiere leyes físicas precisas y que la herencia, un código escrito de algún modo por los cromosomas, debía ser descifrable. Antes de que James Watson y Francis Crick descubrieran el código genético en 1953, Schrödinger describió los rompecabezas de la mutación y la entropía, y señaló que dicha organización se mantiene extrayendo orden del medio ambiente. Muchos de los biólogos con mayor éxito del último siglo leyeron *¿Qué es la vida?* mientras estaban en el instituto o la universidad y afirmaron que cambió el curso de sus vidas.

Con este libro espero incitar a los jóvenes ingenieros y científicos a estudiar la corteza cerebral, adoptar el marco de la memoria-predicción y construir máquinas inteligentes. En su momento culminante, la inteligencia artificial fue un gran movimiento. Tenía revistas, programas de doctorado, libros, planes comerciales y empresarios. De igual modo, las redes neuronales suscitaron una enorme expectación cuando la disciplina floreció en la década de 1980. Pero los marcos científicos subyacentes en la inteligencia artificial y las redes neuronales no eran los adecuados para construir máquinas inteligentes.

Sugiero que ahora contamos con un nuevo camino que seguir más prometedor. Si estás en el instituto o la universidad y este libro te motiva para que te dediques a esta tecnología —construir las primeras máquinas verdaderamente inteligentes, ayudar a poner en marcha una industria—, te animo a que lo hagas. Logremos que suceda. Uno de los trucos del éxito empresarial es que se debe saltar de cabeza el primero a un nuevo campo antes de estar totalmente seguro de que se va a lograr el triunfo. La oportunidad es importante. Si se salta demasiado pronto, toca luchar; si se espera hasta que la incertidumbre desaparece, es demasiado tarde. Creo firmemente que ahora es el momento de comenzar a diseñar y construir sistemas de memoria semejantes a la corteza cerebral. Este campo será importantísimo



científica y comercialmente. Los Intel y Microsoft de la nueva industria de las memorias jerárquicas se pondrán en marcha antes de diez años. Acometer una empresa a esta escala puede resultar arriesgado desde el punto de vista financiero o exigente desde el intelectual, pero siempre merece la pena intentarlo. Espero que os unáis a mí y a muchos otros que aceptan el desafío para crear una de las mayores tecnologías que el mundo ha visto jamás.

## **Apéndice: Predicciones Verificables**

**Toda** teoría debe conducir a predicciones verificables, puesto que el único modo seguro de determinar la validez de una nueva idea es la prueba experimental. Por suerte, el marco de memoria-predicción se basa en la biología y lleva a varias predicciones específicas y nuevas que pueden demostrarse. En este apéndice enumero predicciones capaces de refutar o apoyar las propuestas realizadas en el libro. Este material es algo más avanzado que el presentado en el capítulo 6 y no requiere en absoluto entender el resto del libro. Varias de las predicciones solo pueden realizarse en animales o sujetos humanos despiertos, porque las pruebas suponen expectación y predicción del resultado de un estímulo, lo cual suele requerir el estado de vigilia. Las predicciones no están clasificadas por orden de importancia.

### **Predicción 1**

Debemos encontrar células en todas las áreas de la corteza cerebral, incluida la sensorial primaria, que muestren un aumento de actividad como anticipación a un acontecimiento sensorial y no como reacción a dicho acontecimiento sensorial.

Por ejemplo, Tony Zador, de Cold Spring Harbor Laboratories, ha descubierto células en la corteza cerebral auditiva primaria de las ratas que se estimulan justo cuando la rata espera escuchar un sonido, incluso cuando no lo hay (correspondencia particular). Esta debe ser una propiedad general de la corteza cerebral. Tenemos que encontrar una actividad anticipatoria similar en la corteza cerebral visual y somatosensorial. Las células que se activan como anticipación de una entrada sensorial son la definición de la previsión, premisa básica del marco de memoria-predicción.

### **Predicción 2**

Cuanto más específica en el espacio sea una predicción, más cerca de la corteza cerebral sensorial primaria debemos encontrar células que se activen adelantándose a un acontecimiento.

Si se entrenara a un mono con secuencias de patrones visuales de tal modo que fuera capaz de adelantar un patrón visual particular en un momento preciso, deberíamos encontrar células que muestren un aumento de actividad justo cuando se espera el patrón anticipado (reformulación de la predicción 1). Si el mono aprendiera a ver un rostro pero no supiera con exactitud qué cara aparecería o cómo lo haría, deberíamos esperar encontrar células anticipatorias en las áreas de reconocimiento de rostros, pero no en las áreas visuales inferiores. Sin embargo, si el mono se fija en un objetivo y ha aprendido a esperar un patrón particular en una ubicación precisa de su campo visual, deberíamos encontrar células anticipatorias en V1 o cerca. La actividad que representa la predicción fluye hacia

abajo de la jerarquía cortical lo más lejos que puede según la especificidad de la predicción. A veces logra recorrer todo el camino hasta las áreas sensoriales primarias y otras se detiene en regiones superiores. Deben existir resultados similares en otras modalidades sensoriales.

### **Predicción 3**

Las células que muestran un aumento de actividad como adelanto a una entrada sensorial deberían situarse de modo preferente en las capas corticales 2,3 y 6, y la predicción debería detener su descenso por la jerarquía en las capas 2 y 3.

Las predicciones que se desplazan hacia abajo de la jerarquía cortical lo hacen a través de las células de las capas 2 y 3, que luego se proyectan a la capa 6. Estas células de la capa 6 se proyectan extensamente por la capa 1 de la región de debajo de la jerarquía, activando otro conjunto de células de las capas 2 y 3, y así sucesivamente. Por lo tanto, es en las células de estas capas (2, 3 y 6) donde debemos encontrar actividad anticipatoria. Recordemos que las células activas de las capas 2 y 3 representan un conjunto de posibles columnas activas; son predicciones posibles. Las células activas de la capa 6 representan un número menor de columnas; son las predicciones específicas de una región de la corteza cerebral. Cuando la predicción desciende por la jerarquía, la actividad acabará deteniéndose en las capas 2 y 3. Por ejemplo, pongamos que una rata ha aprendido a anticipar uno de dos tonos auditivos diferentes. Basándose en una pista externa, la rata sabe cuándo escuchará uno de esos dos tonos, pero no es capaz de predecir cuál. En este escenario, debemos esperar ver actividad anticipatoria en las capas 2 o 3 en las columnas que representen ambos tonos. No tiene que haber actividad en la capa 6 de la misma región, porque el animal no puede predecir el tono exacto que escuchará. Si en otra prueba el animal es capaz de predecir el tono exacto que escuchará, debemos ver actividad en la capa 6 en las columnas que responden a ese tono específico.

No podemos descartar por completo la posibilidad de encontrar células anticipatorias en las capas 4 y 5. Por ejemplo, es probable que haya varias clases de células en estas capas con función desconocida. Por lo tanto, esta predicción es relativamente débil, pero sigo creyendo que merece la pena mencionarla.

### **Predicción 4**

Una clase de células de las capas 2 y 3 debería recibir de forma preferente entradas de las células de la capa 6 en regiones corticales superiores.

Parte del modelo de memoria-predicción consiste en que las secuencias aprendidas de patrones que aparecen juntas desarrollan una representación invariable temporalmente constante, lo que denomino un "nombre". Propongo que este nombre es un conjunto de células de las capas 2 o 3 que cruza una región de la

corteza cerebral en diferentes columnas. El conjunto de células permanece activo mientras los acontecimientos que forman parte de la secuencia suceden (por ejemplo, un conjunto de células que permanece activo mientras se escucha una nota de una melodía). Este conjunto de células que representan el nombre de la secuencia es activado mediante realimentación de las células de la capa 6 de las regiones superiores de la corteza cerebral. Sugiero que estas células del nombre eran de la capa 2 debido a su proximidad con la capa 1. Pero podría ser cualquier clase de célula de las capas 2 y 3 que tengan dendritas en la capa 1. Para que funcione el sistema de nombres las dendritas apicales de esas mismas células de nombre tienen que formar sinapsis preferentes con los axones de la capa 1 que se originan en la capa 6 de regiones superiores. Deben evitar formar sinapsis con axones de la capa 1 que se originan en el tálamo. Así pues, la teoría sugiere que hemos de encontrar una clase de células dentro de las capas 2 y 3 con dendritas apicales en la capa 1 que muestren una preferencia manifiesta por formar sinapsis con los axones de las células de la capa 6 de la región superior. Otras células con sinapsis en la capa 1 no deben presentar esta preferencia. Hasta donde alcanzo, esta es una predicción sólida y novedosa.

Una predicción corolario es que debemos encontrar otra clase de células en las capas 2 o 3 cuyas dendritas apicales formen sinapsis preferentes con los axones que se originan en regiones no específicas del tálamo. Estas células predicen elementos siguientes en una secuencia.

### **Predicción 5**

Un conjunto de células de "nombre" descrito en la predicción 4 debería permanecer activo durante las secuencias aprendidas.

Un conjunto de células que permanece activo durante una secuencia aprendida es la definición de un "nombre" para una secuencia predecible. Por lo tanto, deberíamos encontrar conjuntos de células que permanezcan activas incluso cuando la actividad de las células en el resto de una columna (células de las capas 4, 5 y 6) esté cambiando. Por desgracia, no podemos precisar cómo será la actividad de las células de nombre. Por ejemplo, la actividad constante de un patrón de nombre podría ser algo tan sencillo como un único impulso eléctrico que cruzara de forma conjunta todas las células de nombre. Así pues, este grupo de células activas podría ser difícil de detectar.

### **Predicción 6**

Otra clase de células de las capas 2 y 3 (diferente de las células de nombre a las que hemos hecho referencia en las predicciones 4 y 5) deberían activarse en respuesta a una entrada imprevista, pero deberían permanecer inactivas en respuesta a una entrada adelantada.

La idea que se encuentra tras esta predicción es que los acontecimientos no anticipados deben ascender por la jerarquía cortical, pero cuando un acontecimiento sí es anticipado no queremos pasarlo hacia arriba de la jerarquía debido a que se predijo localmente. Por lo tanto, debe haber una clase de células en las capas 2 y 3, diferente de la clase de nombre descrita en las predicciones 4 y 5, que muestra actividad cuando ocurre un hecho imprevisto, pero no cuando el hecho sí se había previsto. Los axones de estas células deben proyectarse a regiones superiores de la corteza cerebral. Propongo un mecanismo para lograr este cambio de actividad. Dicha célula podría inhibirse mediante una interneurona activada por una célula de nombre, pero en este punto no hay modo de establecer una predicción sólida del mecanismo. Todo lo que cabe afirmar es que algunas células deben mostrar esta actividad diferencial. Esta es otra predicción consistente y, hasta donde sé, novedosa.

### **Predicción 7**

En conexión con la predicción 6, los acontecimientos no anticipados deberían propagarse ascendiendo por la jerarquía. Cuanto más novedoso resulte el acontecimiento, más arriba debería fluir la entrada no prevista. Los acontecimientos completamente nuevos tendrían que llegar hasta el hipocampo.

Los patrones bien aprendidos se predicen más abajo en la jerarquía y, por el contrario, cuanto más novedosa es una entrada, hasta más arriba de la jerarquía debe propagarse. Se tiene que diseñar un experimento para captar esta diferencia. Por ejemplo, un ser humano podría escuchar una melodía desconocida pero sencilla. Si el sujeto escucha una nota que, aunque inesperada, es coherente con el estilo de música, dicha nota inesperada debe causar cambios de actividad en la corteza cerebral auditiva, ascendiendo hasta cierto nivel de la jerarquía cortical. Sin embargo, si en lugar de una nota coherente con el estilo de la música el sujeto escucha un sonido sin sentido alguno, como un estallido, esperaríamos cambios de actividad de este sonido para viajar más arriba de la jerarquía cortical. Los resultados deberían cambiar si el sujeto estaba esperando escuchar el estallido, mas en su lugar escuchó la nota. Se podría probar esta predicción con FMRI (imagen de resonancia magnética funcional) en sujetos humanos.

### **Predicción 8**

La comprensión repentina debe dar como resultado una cascada precisa de actividad predictiva que fluye hacia abajo de la jerarquía cortical.

El momento de percepción en el que un patrón sensorial desconcertante acaba entendiéndose —como el reconocimiento de la silueta del perro dalmata de la figura 12— comienza cuando una región de la corteza cerebral intenta una nueva correspondencia de su entrada con la memoria. Si la correspondencia encaja en la región local, las predicciones se pasan hacia abajo de la jerarquía cortical en rápida

sucesión a todas las regiones inferiores. Si se trata de una interpretación correcta del estímulo, cada región de la jerarquía establecerá una predicción acertada en rápida sucesión. El mismo efecto debe ocurrir mientras se ve una imagen con dos interpretaciones, como una silueta en un jarrón que puede parecer dos caras o un cubo de Necker (una imagen de un cubo que aparece de forma alternativa en dos orientaciones diferentes). Cada vez que la percepción de dicha imagen cambia, debemos ver una propagación de nuevas predicciones fluir hacia abajo de la jerarquía. En los niveles inferiores, digamos V1, una columna que representa un segmento de línea de la imagen debe permanecer activo en cualquier percepción de la imagen (suponiendo que los ojos no se hayan movido). Sin embargo, podríamos ver a algunas células de esa columna trocar estados activos. Es decir, el mismo rasgo de bajo nivel existe en cada imagen, pero puede que se activen diferentes células dentro de una columna según las interpretaciones diferentes. El punto principal es que tenemos que ver descender la propagación de las predicciones por la jerarquía cortical cuando cambia una percepción en un nivel superior.

Debe ocurrir una propagación de la predicción similar con cada sacudida visual sobre un objeto determinado.

## **Predicción 9**

El marco de memoria-predicción requiere que las neuronas piramidales puedan detectar coincidencias precisas de entrada sináptica en dendritas finas.

Durante muchos años se pensó que las neuronas podrían ser integradoras que compendian las entradas de todas sus sinapsis para determinar si la neurona debe transmitir un impulso nervioso. En la neurociencia actual hay mucha incertidumbre acerca del comportamiento de las neuronas. Algunas personas siguen manteniendo la idea de que son simples integradoras, y muchos modelos de redes neuronales se construyen con neuronas que funcionan de ese modo. Asimismo, existen muchos modelos que asumen que una neurona se comporta como si cada sección dendrítica operara de forma independiente. El modelo de memoria-predicción requiere que las neuronas sean capaces de detectar la coincidencia de unas cuantas sinapsis activas en un lapso de tiempo reducido. El modelo podría funcionar incluso con una única sinapsis potenciada que bastara para provocar que una célula se estimulara, pero habría más probabilidades si existieran dos o más sinapsis activas próximas en una dendrita fina. De este modo, una neurona con cientos de sinapsis puede aprender a estimular muchos patrones de entrada diferentes precisos y separados. No se trata de una idea nueva, y hay pruebas que la apoyan. Sin embargo, supone un alejamiento radical del modelo estándar empleado durante muchos años. Si se demostrara que las neuronas no se estimulan con patrones de entrada precisos y dispersos, sería difícil mantener intacta la teoría de la memoria-predicción. Las sinapsis en las dendritas gruesas del

cuerpo celular o cercanas a este no necesitan funcionar de esta forma; solo las múltiples sinapsis de las dendritas finas.

### **Predicción 10**

La extensión de árboles axonales y dendríticos debería ampliarse a medida que ascendemos en la jerarquía cortical y nos alejamos de la corteza cerebral sensorial primaria.

He sostenido que V1 y otras regiones sensoriales primarias no son grandes, sino que en realidad están compuestas por muchas regiones pequeñas. Estas subregiones se encuentran probablemente entre las más pequeñas de la corteza cerebral. Esta interpretación de V1 y otras regiones sensoriales ayuda a comprender cómo funciona la jerarquía, aunque no contradice mucho de lo que creemos sobre dichas áreas sensoriales primarias. Una manifestación física de que existen regiones pequeñas es que la expansión de los árboles axonales y dendríticos debe estrecharse en la corteza cerebral sensorial primaria y agrandarse a medida que se asciende de V1 a V2 y V4, por ejemplo. Básicamente, esta ampliación refleja el tamaño de una región. Ya existen algunas pruebas experimentales que apoyan esta idea, así que no es una predicción novedosa. Sin embargo, debemos ver esta ampliación de los árboles en la corteza cerebral visual, somatosensorial y motora, y es probable que también en la auditiva.

### **Predicción 11**

Las representaciones descienden por la jerarquía con el entrenamiento.

He sostenido que mediante el entrenamiento repetido la corteza cerebral aprendería secuencias en regiones inferiores de la jerarquía cortical, lo cual se deduce de forma natural del modo como la memoria de las secuencias de patrones cambia el patrón de entrada pasado a las regiones superiores siguientes de la corteza cerebral. Este proceso tiene un par de consecuencias. Una es que debemos encontrar células que respondan a un estímulo complejo más abajo de la corteza cerebral tras un entrenamiento extenso y más arriba tras un entrenamiento moderado. En un ser humano, por ejemplo, esperaríamos encontrar células que respondan a las letras impresas en una región como IT tras un entrenamiento para reconocer letras particulares. Pero después de aprender a leer palabras completas, yo esperaría encontrar células que respondieran a letras en partes diferentes de V4, además de IT. Resultados similares deben obtenerse con otras especies, regiones y estímulos. Otra consecuencia de este proceso de aprendizaje es que debe cambiar el lugar donde se detectan recuerdos y errores. Es decir, las sensaciones de patrones bien aprendidos deben propagarse a menos distancia de ascenso por la jerarquía, lo cual podría detectarse mediante técnicas de imagen. También tendríamos que ser capaces de detectar un cambio en los

tiempos de reacción ante ciertos estímulos, porque las entradas no tendrán que viajar tan lejos en la corteza cerebral para ser reconocidas y recordadas.

## **Predicción 12**

Las representaciones invariables deben encontrarse en todas las áreas corticales.

Es bien sabido que existen células que responden a entradas muy selectivas invariables a los detalles. Se han observado células que responden a rostros, manos, Bill Clinton, etc. El modelo de memoria-predicción vaticina que todas las regiones de la corteza cerebral deben formar representaciones invariables. Estas deben reflejar todas las modalidades sensoriales bajo una región de la corteza cerebral. Por ejemplo, si yo tuviera una célula Bill Clinton en la corteza cerebral visual, se estimularía cada vez que yo viera a Bill Clinton. Si yo tuviera una célula Bill Clinton en la corteza cerebral auditiva, se estimularía cada vez que yo escuchara el nombre "Bill Clinton". Luego esperaríamos encontrar células en áreas de asociación que recibieran una entrada tanto visual como auditiva que respondiera a la vista o el nombre pronunciado de Bill Clinton. Deberíamos encontrar representaciones invariables en todas las modalidades sensoriales e incluso en la corteza cerebral motora. En esta las células representarían secuencias motoras complejas. Cuanto más arriba en la jerarquía motora, más complejas e invariables deberían ser las representaciones. (Estudios recientes parecen haber descubierto células que activan complejos movimientos de la mano a la boca en los monos.) No se trata de predicciones nuevas. La mayoría de los investigadores creen en la idea general de que las representaciones invariables se forman en muchas localizaciones por toda la corteza cerebral. Sin embargo, aunque lo he analizado como un hecho, no se ha demostrado en todas partes. El modelo de memoria-predicción presiente que veremos dichas células en cada una de las regiones de la corteza cerebral.

\* \* \*

Las predicciones anteriores son algunas de las formas de probar el modelo de este libro. Estoy seguro de que hay otras. Sin embargo, no es posible demostrar si una teoría es acertada; solo si es errónea. Así pues, aunque se demostrara que todas las predicciones que acabo de enumerar son ciertas, no constituirían una prueba de que la hipótesis de memoria-predicción es acertada, pero sí sería una prueba sólida en apoyo de la teoría. Lo contrario también es cierto. Si algunas de las predicciones precedentes resultan no ser verdaderas, no invalidarían necesariamente la tesis completa. Para algunas de las predicciones hay modos alternativos de alcanzar la conducta necesaria. Por ejemplo, hay otros modos de crear secuencias de nombres. Este apéndice solo pretende mostrar que el modelo conduce a varias predicciones y, por lo tanto, puede ser probado. Diseñar experimentos es una labor exigente y necesitaría mucho más análisis del que resulta adecuado para este libro. También sería excelente encontrar modos de



probar esta teoría con técnicas de imagen como FMRI. Existen muchos grandes laboratorios especializados y dichos experimentos pueden realizarse con relativa rapidez comparado con el registro directo de las células.

## **Bibliografía**

**La** mayoría de los libros científicos y artículos de revistas cuentan con extensas bibliografías que sirven tanto para catalogar las contribuciones de otros como para ayudar al lector. Como este libro pretende llegar a un público variado en el que se incluyen personas sin un conocimiento previo sobre neurociencia, hemos evitado escribir con estilo académico. De igual modo, esta bibliografía se ha diseñado ante todo para que sirva de ayuda al lector no experto que desee aprender más. No enumero toda la investigación importante publicada, ni intento citar a todas las personas que han hecho descubrimientos trascendentales en este campo, sino que me limito a presentar una selección de obras que creo que serían buenos materiales para que un lector interesado aprendiera más sobre los cerebros. También incluyo algunos textos que me resultaron útiles, pero que en su mayoría están dirigidos al especialista. Pueden encontrar análisis más profundos sobre muchos de estos temas en Internet.

Por desgracia, solo encontrarán unas cuantas referencias a teorías generales sobre el cerebro, porque, como he señalado en el prólogo, no se ha escrito mucho sobre este tema, e incluso menos sobre las propuestas específicas de este libro.

## **Historia de la Inteligencia Artificial y las Redes Neuronales**

Arbib,

Michael A., ed., *The Handbook of Brain Theory and Neural Networks*, 2.<sup>a</sup> ed. (Cambridge, Mass.: MIT Press, 2003).

Existen muchos libros sobre las redes neuronales, pero la mayoría no resultan muy útiles para comprender el funcionamiento del cerebro. Este es una recopilación de artículos cortos sobre varios de los diferentes temas relacionados con las redes neuronales. Su lectura proporciona una visión general del campo.

Baumgartner,

Peter, y Sabine Payr, eds., *Speaking Minds: Interview with Twenty Eminent Cognitive Scientists* (Princeton, N. J.: Princeton University Press, 1995).

Este libro contiene entrevistas interesantes con muchos de los pensadores más importantes de la inteligencia artificial, las redes neuronales y la ciencia cognitiva. Es una sinopsis sencilla y agradable de la historia reciente y el espíritu del pensamiento sobre la inteligencia.

Dreyfus,

Hubert L., *What Computers Still Can't Do: A Critique of Artificial Reason* (Cambridge, Mass.: MIT Press, 1992).

Una dura crítica de la inteligencia artificial originalmente publicada como *What Computers Can't Do* y reeditada años después con el título revisado. Se trata de una historia en profundidad de la inteligencia artificial escrita por uno de sus más acérrimos críticos.

Hebb,

D. O., *The Organization of Behavior: A Neuropsychological Theory* (Nueva York: Wiley, 1949).

Escrito en 1949, es el texto clásico en el que Hebb propuso la regla de aprendizaje "hebbiano" ahora aceptada. A pesar de ser antiguo, sigue conteniendo algunas ideas interesantes y resulta importante sobre todo desde la perspectiva histórica.

Kohonen,

Teuvo, *Self-organization and Associative Memories* (Nueva York: Springer Verlag, 1984).

Para entender cómo funciona la corteza cerebral y almacena secuencias de patrones, sirve de ayuda conocer las memorias autoasociativas. Aunque se ha escrito mucho al respecto, no he encontrado fuentes impresas que presenten un resumen legible de lo que considero importante. Kohonen es un pionero en este campo. Este libro es difícil de conseguir y no demasiado sencillo de leer, pero aborda lo básico de las memorias autoasociativas, incluida la memoria de secuencias.

Mcculloch,

W. S., y W. Pltts, "A Logical Calculus of the Ideas Immanent in Nervous Activity", *Bulletin of Mathematical Biophysics*, vol. 5 (1943), págs. 115-133.

Artículo clásico de este autor donde propone que las neuronas pueden concebirse en esencia como puertas lógicas. Lo analizo en el capítulo 1.

Searle,

J. R., "Minds, Brains, and Programs", *The Behavioral and Brain Sciences*, vol. 3 (1980), págs. 417-424.

Presenta el famoso argumento de la "habitación china" contra la computación como modelo para la mente. Se pueden encontrar muchas descripciones y análisis del experimento mental de Searle en Internet.

Turing,

A. M., "Computing Machinery and Intelligence", *Mind*, vol. 5 (1950), págs. 433-460.

Presenta el famoso "test de Turing" para detectar la presencia de inteligencia. De nuevo, se pueden encontrar muchas referencias y análisis sobre dicho test en Internet.

### **Corteza Cerebral y Neurociencia General**

Los libros que siguen se recomiendan para quienes deseen saber más sobre neurobiología y la corteza cerebral.

Crick,

Francis H. C., "Thinking about the Brain", *Scientific American*, vol. 241 (septiembre 1979), págs. 181-188. También disponible en *The Brain: A Scientific American Book* (San Francisco: W. H. Freeman, 1979).

Este es el artículo que hizo que me interesara en los cerebros. Aunque tiene una antigüedad de veinticinco años, me sigue resultando inspirador.

Koch,

Christof, *Quest for Consciousness: A Neurobiological Approach* (Denver, Colo.: Roberts and Co., 2004).

Todos los años se publican varios libros sobre el cerebro de interés general. Este de Christof Koch trata de la conciencia, pero aborda la mayoría de los temas importantes sobre cerebros, neuroanatomía, neurofisiología y conciencia. Si desea una introducción básica a la neurobiología y la ciencia cerebral en un único libro de fácil lectura, este sería uno bueno para comenzar.

Mountcastle,

Vernon B., *Perceptual Neuroscience: The Cerebral Cortex* (Cambridge, Mass.: Harvard University Press, 1998).

Un gran libro dedicado por completo a la corteza cerebral. Está muy bien escrito, tiene una división clara y, aunque es técnico, su lectura resulta agradable. Es una de las mejores introducciones a la corteza cerebral.

Kandel,

Eric R.; James H. Schwartz y Thomas M. Jessell, eds., *Principles of Neural Science*, 4.<sup>a</sup> ed. (Nueva York: McGraw-Hill, 2000).

Se trata de una enciclopedia de un solo volumen sobre la neurociencia. Este libro enorme no es lectura de cabecera, pero sí una buena obra de referencia con que contar. Proporciona introducciones detalladas a todas las partes del sistema nervioso, incluidas neuronas, órganos sensoriales, neurotransmisores y demás.

Shepherd,

Gordon M., ed., *The Synaptic Organization of the Brain*, 5.<sup>a</sup> ed. (Nueva York: Oxford University Press, 2004).

Este libro me ha sido útil, aunque prefería las ediciones anteriores, escritas por un solo autor. Es una fuente técnica sobre todas las partes del cerebro, en particular las sinapsis. Lo uso como referencia.

Koch,

Christof, y Joel L. Davis, eds., *Large-scale Neuronal Theories of the Brain* (Cambridge, Mass.: MIT Press, 1994).

Se ha escrito muy poco sobre las teorías generales del cerebro. Este libro es una compilación de artículos sobre este tema, aunque la mayoría de los que aparecen no cumplen con la meta sugerida en el título. Proporciona una visión general de los planteamientos diversos que se están adoptando para comprender cómo funciona el cerebro en general. Como en muchos textos recientes, se pueden encontrar elementos del marco de memoria-predicción a lo largo del libro.

Braitenberg,

Valentino, y Almut Schüz, *Cortex: Statistics and Geometry of Neuronal Connectivity*, 2.<sup>a</sup> ed. (Nueva York: Springer Verlag, 1998).

Este libro describe las propiedades estadísticas del cerebro del ratón. Soy consciente de que no parece muy apasionante, pero se trata de un libro alentador y útil. Cuenta la historia de la corteza cerebral en números.

### **Artículos Específicos Sobre Neurociencia**

Los artículos siguientes constituyen las fuentes originales de algunos de los conceptos importantes que se describen en este libro. La mayoría solo pueden encontrarse en una biblioteca universitaria *online*.

Mountcastle,

Vernon B., "An Organizing Principle for Cerebral Function: The Unit Model and the Distributed System", en Gerald M. Edelman y Vernon B. Mountcastle, eds., *The Mindful Brain* (Cambridge, Mass.: MIT Press, 1978).

En este artículo leí por primera vez la propuesta de Mountcastle de que toda la corteza cerebral funciona según un principio común. También propone que la columna cortical es la unidad básica de computación. Estas ideas son la premisa e inspiración de la teoría presentada en este libro.

Felleman,

D. J., y D. C. Van Essen, "Distributed Hierarchical Processing in the Primate Cerebral Cortex", *Cerebral Cortex*, vol. 1 (enero-febrero 1991), págs. 1-47.

Este es el artículo ya clásico que describe la organización jerárquica de la corteza cerebral visual. El marco de memoria-predicción se basa en el supuesto de que no solo el sistema visual, sino toda la corteza cerebral, posee una estructura jerárquica.

Sherman,

S. M., y R. W. Guillery, "The Role of the Thalamus in the Flow of Information to the Cortex", en *Philosophical Transactions of the Royal Society of London*, vol. 357, núm. 1.428 (2002), págs. 1695-1708.

Proporciona una visión general de la organización del tálamo y plantea la hipótesis de Sherman-Guillery según la cual el tálamo sirve para regular el flujo de

información entre áreas corticales. En el capítulo 6 desarrollo esta idea en el epígrafe titulado "Una ruta alternativa para ascender en la jerarquía".

Rao,

R. P, y D. H. Ballard, "Predictive Coding in the Visual Cortex: A Functional Interpretation of Some Extra-Classical Receptive-field Effects", *Nature Neuroscience*, vol. 2, núm. 1 (1999), págs. 78-87.

Incluyo este artículo como ejemplo de la investigación reciente que habla de predicción y jerarquías. Presenta un modelo de realimentación en las jerarquías corticales en el que las neuronas de las áreas superiores tratan de predecir patrones de actividad en las áreas inferiores.

## Agradecimientos

**Cuando** alguien me pregunta cómo me gano la vida, nunca sé qué responder. La verdad es que me la gano haciendo muy poco, pero me he rodeado de gente que sí parece hacer muchas cosas. Mi contribución es llamarles la atención de cuando en cuando y tratar de redirigir al equipo por un nuevo camino cuando resulta necesario. Si he obtenido algún éxito en mi carrera se lo debo en buena parte al trabajo duro y a la inteligencia de mis colegas.

He tenido la suerte de conocer a muchos científicos y casi todos me han enseñado algo; por lo tanto, todos ellos han contribuido a las ideas que aparecen en este libro. Les doy las gracias a todos, pero solo puedo mencionar aquí a unos cuantos. Bruno Olshausen, que trabaja en el Redwood Neuroscience Institute (RNI) y la Universidad de California, en Davis, es una enciclopedia andante sobre neurociencia; me señala de forma constante lo que no sé y me sugiere formas de rectificar mi ignorancia, lo cual es una de las cosas más valiosas que puede hacer alguien. Bill Softky, también del RNI, fue la primera persona que me ilustró sobre la reducción del tiempo en la jerarquía cortical y las propiedades de las dendritas finas. Rick Granger, de la Universidad de California, en Irvine, me dio la oportunidad de comprender la memoria de secuencias y el papel que podía desempeñar el tálamo. Bob Knight, de la Universidad de California, en Berkeley, y Christof Koch, del California Institute of Technology, han sido fundamentales para la formación del Redwood Neuroscience Institute y muchos otros asuntos científicos. Todo el personal del RNI me ha puesto a prueba y obligado a perfeccionar mis ideas; muchas de las propuestas de este libro fueron el resultado directo de reuniones y discusiones celebradas allí. Gracias a todos.

Donna Dubinsky y Ed Colligan han sido mis socios empresariales durante una docena de años. Por su buen trabajo y ayuda logré ser empresario mientras trabajaba a media jornada sobre la teoría del cerebro, algo poco habitual. Donna solía afirmar que uno de sus objetivos era que nuestra empresa tuviera éxito para que así yo pudiera dedicar más tiempo a la teoría del cerebro. Este libro no existiría si no fuera por Donna y Ed.

No podría haber escrito Sobre la inteligencia sin multitud de ayudas. Jim Levine, mi agente, creyó en el texto incluso antes de que yo supiera lo que iba a escribir. No hagan un libro sin un agente como Jim. Me presentó a Sandra Blakeslee, mi coautora. Yo quería que este libro resultara legible para un público amplio, y Sandy fue esencial para lograrlo. Yo soy el culpable de los epígrafes difíciles. Matthew Blakeslee, hijo de Sandy y también escritor científico, proporcionó varios ejemplos usados en el texto y sugirió el término marco de memoria-predicción. Ha sido muy agradable trabajar con todo el personal de Henry Holt. Me gustaría dar las gracias en especial a John Sterling, director y editor de dicha editorial. Solo lo vi una vez cara a cara y hablamos por teléfono unas cuantas veces, pero no le hizo falta más para tener una enorme repercusión sobre la estructura del libro. Comprendió de



inmediato los temas a los que me enfrentaría al proponer una teoría sobre la inteligencia, y luego sugirió cómo debía escribirse y disponerse el texto.

Quiero agradecerles a mis hijas, Anne y Kate, que no se hayan quejado cuando su papá pasaba muchos fines de semana ante el teclado del ordenador. Y para finalizar, doy las gracias a mi esposa, Janet. Estar casada conmigo no es nada fácil. La quiero a ella más que a los cerebros.

## **Sobre los Autores**

JEFF HAWKINS es uno de los ingenieros informáticos y empresarios de Silicon Valley que ha alcanzado mayor éxito y reconocimiento. Fundador de Palm Computing, Handspring, y el Redwood Neuroscience Institute, creado para fomentar la investigación sobre la memoria y la cognición, es asimismo miembro del consejo científico de Cold Spring Harbor Laboratories.

SANDRA BLAKESKEE lleva más de treinta años escribiendo sobre ciencia y medicina para *The New York Times* y es co-autora de *Phantoms in the Brain* y de los exitosos libros sobre psicología y matrimonio de Judith Wallerstein.

ÿ