



La **fábrica** de los **sueños**

Pedro Miramontes



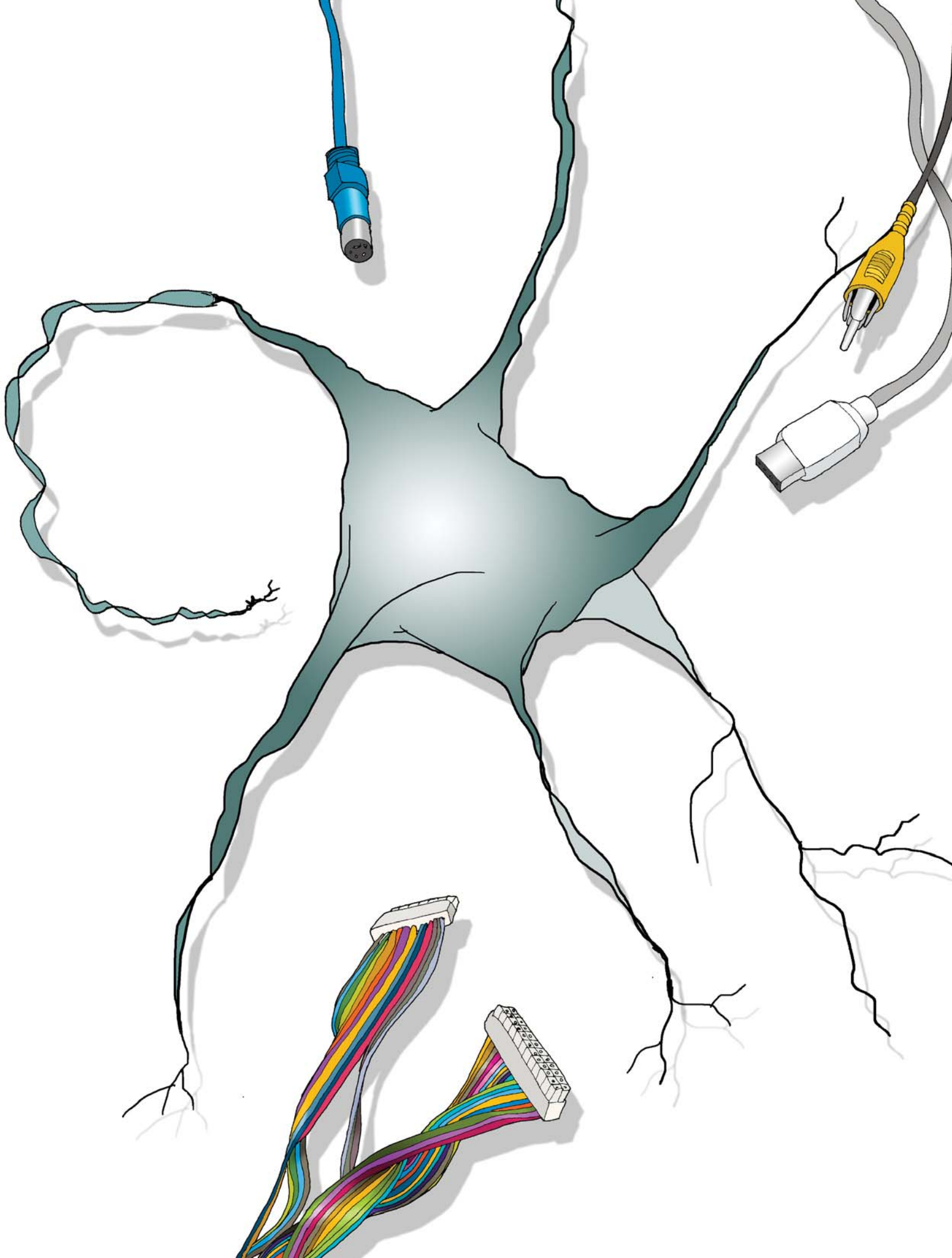
A Alan Mathison Turing se le reconocen grandes logros en los campos de la teoría de la computación, la criptografía, la morfogénesis y la inteligencia artificial. Lamentablemente, sus aportaciones al cómputo neuronal permanecen en el olvido. Este ensayo destaca algunos de los puntos principales de un manuscrito oscuro y desconocido que escribió en 1948 con el título “Máquinas inteligentes”.

El mapa de la mente

Cavendish, Vermont, Estados Unidos. El 13 de septiembre de 1848, Phineas Gage llegó como todas las mañanas a su trabajo en esta localidad de la Costa Este. Como capataz en la construcción del terraplén del ferrocarril, tenía la obligación de supervisar la colocación de las cargas de dinamita empleadas para remover material y preparar el nivelado del terreno. Por negligencia o descuido, ese día Gage entró a la historia como uno de los casos más extremos que se conocen de supervivencia a un accidente, que le provocó un daño cerebral extremo.

La explosión de un cartucho de dinamita lanzó una barra de metal de un metro de largo por cuatro centímetros de diámetro a través del rostro de Gage (Figura 1), penetrando de abajo hacia arriba por la mejilla izquierda para atravesar el lóbulo frontal del cerebro, salir por la región occipital y caer luego a varios metros de distancia. La lesión fue espantosa, la barra destrozó buena parte del cerebro de Gage y el daño producido por los fragmentos de hueso, que interesaron prácticamente todo el interior del encéfalo, fue mayúsculo. De manera asombrosa, Gage no murió. De hecho, se recuperó totalmente; a los tres meses fue dado de alta y volvió al trabajo.

A partir de ese infausto día, Phineas Gage se volvió un ejemplo arquetípico que ilustra la robustez del cerebro y que, en su época, se usó como argumento en un debate que dominaba los círculos de la psicología: la especialización o locali-



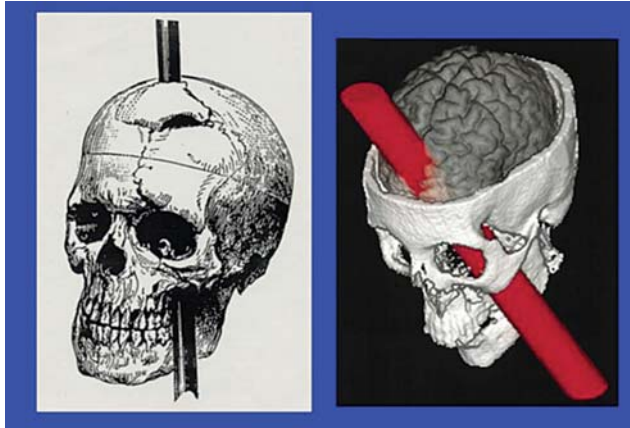


Figura 1. El cráneo de Phineas Gage después del accidente, según un grabado de la época, y una reconstrucción moderna.

zación de la función de las regiones cerebrales. La escuela fundada por Franz Joseph Gall afirmaba que existía una correspondencia biunívoca entre las regiones cerebrales y las funciones cognitivas, de comportamiento y sensoriales del cuerpo entero. Según su teoría –la frenología–, a partir de la forma y volumen de las regiones de la corteza cerebral se podía inferir no únicamente el estado de salud del individuo, sino su personalidad y sus potencialidades intelectuales y sociales (Figura 2).

El caso de Gage fue un elemento en contra de la postura que defendía la localización de las funciones corporales en regiones del cerebro. El debate continuó, sin embargo, pues si bien Gage vivió 16 años después del accidente, su personalidad cambió de manera notable. Sus biógrafos lo describen como sereno, sociable y bien comportado antes del accidente, e impulsivo, soez e irreverente después. De cualquier modo, aunque muy influyente en su época, ahora la frenología está completamente desacreditada, se le considera una *seudociencia*. Sin embargo, la historia de Gage deja en el aire algunas interrogantes: ¿de qué están hechos el pensamiento, el intelecto, la personalidad y la conciencia de las personas? ¿Cómo funciona el cerebro, que aun después de perder una porción considerable de su masa puede mantener sin cambio la mayoría de sus funciones?

Leonardo da Vinci equiparaba el cuerpo humano con una maquinaria donde los órganos, músculos, huesos y tendones jugaban el papel de palancas, resortes,

engranes y bombas. Esta visión, bajo el nombre de *meccanicismo*, permaneció vigente durante mucho tiempo. René Descartes separó mente de cuerpo y propuso un *dualismo* en el cual el cuerpo actuaba como una maquinaria y la mente no se encontraba sujeta a ninguna ley de la materia. Un problema muy serio era que, si bien ya se sabía que el cerebro controla muchas funciones somáticas, el alma tendría que estar ligada, de algún modo, al cuerpo. Descartes propuso, sin evidencia experimental o justificación lógica, que el asiento del alma era la glándula pineal, un pequeño órgano que reside en el paleoencéfalo y que ahora se sabe que emite la hormona melatonina en cantidades que varían según las fases de luz u oscuridad que experimente el individuo. Esta glándula, según Descartes, sería una especie de interfaz entre cuerpo y mente. Son bien conocidos los problemas del esquema cartesiano y no los discutiré aquí: Baruch Spinoza ya lo puso en su lugar.

Desde la imperial capital de Prusia, Königsberg, hoy Kaliningrado en Rusia, Immanuel Kant fue el primer

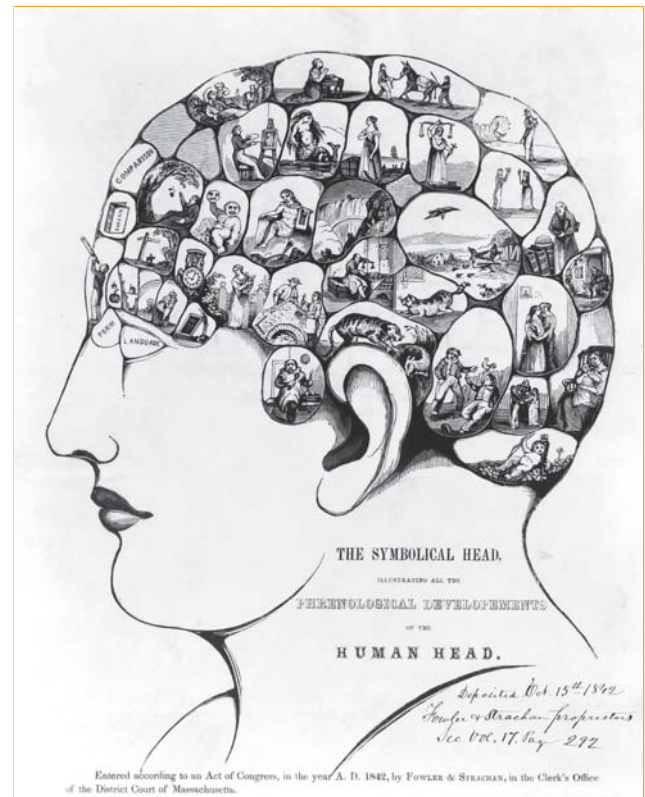


Figura 2. Mapa de las regiones del cerebro. En cada una aparece una alegoría del comportamiento, virtud o vicio que se suponía localizado en dicha región.

"Intelligent Machinery".

I propose to investigate the question as to whether it is possible for machinery to show intelligent behaviour. It is usually assumed without argument that it is not possible. Common catch phrases such as 'acting like a machine', 'purely mechanical behaviour' reveal this common attitude. It is not difficult to see why such an attitude should have arisen. Some of the reasons are

Figura 3. El título y primer párrafo del ensayo de Turing de 1948. Ésta y las siguientes figuras provienen del manuscrito de Turing (tomado de www.alanturing.net, febrero de 2013).

pensador en distinguir claramente entre un *mecanismo* y un *organismo*. En el primero, cada parte tiene una función específica que se puede definir incluso aislando la parte del total. Por ejemplo, la brújula tiene una función bien definida en una nave; pero un barco sin brújula sigue siendo un barco, mientras que un humano sin hígado no lo es. En un organismo no hay “partes” como se entiende habitualmente; lo que hay son subsistemas que tienen un alto grado de integración con el todo y existe una interdependencia esencial en la cual las “partes” se influyen mutuamente, en mayor o menor grado, unas con otras.

Para los años treinta y cuarenta del siglo XX ya se sabía que las funciones neurológicas, y en particular las cognitivas y sensoriales, se encontraban, de alguna manera, en la masa encefálica. No puede ser de otra manera: o están ahí o se vuelve a la separación cuerpo-alma, dejando a esta última bajo la supervisión de un creador.

En este contexto, en 1948, Alan Turing escribió un manuscrito llamado “Máquinas inteligentes” en el que propone la existencia de dispositivos formados a partir de componentes unitarios que tuviesen la capacidad de conmutar entre dos o más estados, y de comunicar su estado al resto del sistema; es decir, Turing se imaginó que, en analogía con la corteza cerebral, podría darse la existencia de máquinas con capacidades como las mencionadas arriba (Figura 3).

A lo largo del manuscrito, Turing desarrolla la propuesta de una máquina que sería capaz de reproducir rasgos que nosotros calificamos de “inteligentes”. Con gran visión, se encargó de pulir los detalles más relevantes de su funcionamiento, e incluso llegó a confiar en que tal máquina podría llegar un día a ser construida.

El borrador de Turing jamás vio la luz. Su jefe en el *National Physical Laboratory* era en ese momento Sir Charles Darwin, nieto del famoso naturalista inglés, quien descalificó el manuscrito de Turing catalogándolo como “un ensayo de estudiante” y pidiéndole que se dedicase a labores más serias.

● Neuronas artificiales

En la Figura 4 se muestra el dibujo, de la pluma de Santiago Ramón y Cajal, de una neurona. Aunque existen muchos tipos distintos de neuronas en el sistema nervioso, todas (o casi todas) comparten los rasgos generales que son relevantes para su funcionamiento: un cuerpo celular, llamado *soma*, donde se llevan a cabo las funciones metabólicas básicas de la célula, y se recolectan las señales eléctricas que entran por las fibras. En el soma se procesa la información que entra, y la respuesta de salida se canaliza por una vía llamada *axón*, que a su vez se puede ramificar y conectarse con *dendritas* que irán a otra neurona. En un mamífero típico, cada neurona tiene alrededor de miles o decenas de miles de conexiones de entrada y salida.

En el sistema nervioso central, las señales que viajan por los axones son *potenciales de acción*: picos que viajan sin perder su forma y amplitud. El potencial de acción llega eventualmente a las *juntas sinápticas*, donde se liberan moléculas de neurotransmisores y se generan *potenciales postsinápticos* que se difunden a través de las dendritas hacia el soma.

En 1943, los estadounidenses McCulloch y Pitts propusieron un modelo matemático de la neurona biológica (Figura 5), trabajo que al parecer Turing

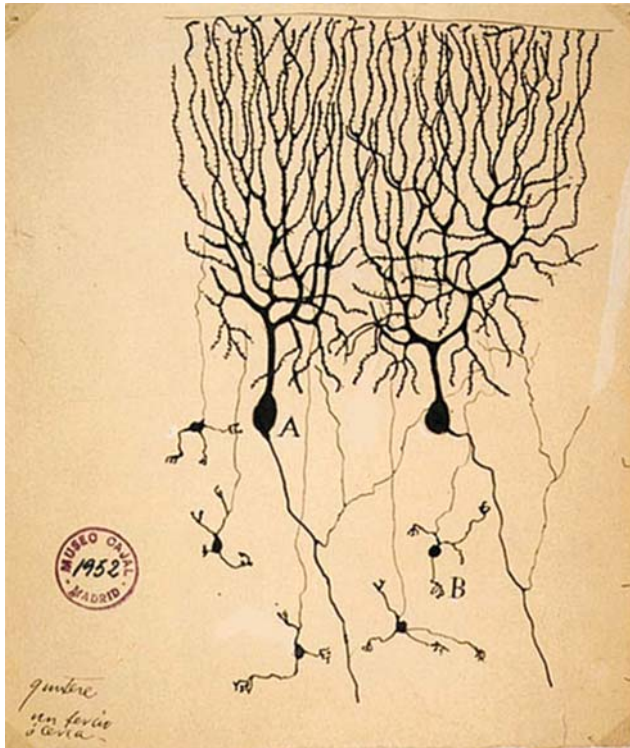


Figura 4. Una neurona de Purkinje, en dibujo autógrafo de Santiago Ramón y Cajal. Éstas son las neuronas más grandes de la corteza cerebral.

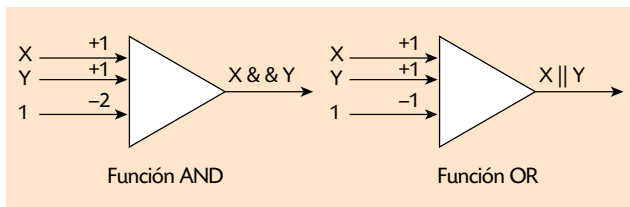


Figura 5. La neurona matemática de McCullough y Pitts puede implementar funciones lógicas.

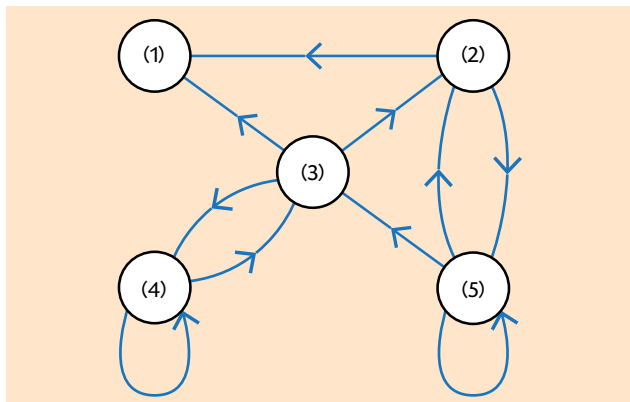


Figura 6. Una máquina desorganizada del tipo A. Obsérvese que cada nodo o neurona tiene dos entradas.

desconocía, por lo que podemos considerar que la propuesta de neurona matemática de Turing, aunque no es pionera, si es original.

Turing llamó a su primer modelo de máquina inteligente “máquina desorganizada del tipo A”. Las máquinas de tipo A son redes de nodos binarios conectadas entre sí al azar (Figura 6). Cada nodo (o neurona) puede encontrarse en el estado cero o uno; recibe la información binaria proveniente de otras neuronas a través de dos canales y ejecuta la operación NAND (la función lógica *Not And*). Es razonable suponer que Turing sabía que cualquier operador binario bivaluado se puede escribir a partir de una serie de combinaciones de operadores NAND. La máquina de tipo A es entonces un caso particular de las redes booleanas modernas.

En la Tabla 1 se puede apreciar que las cinco neuronas reciben entradas de sus vecinas alimentadoras:

Tabla 1.

r	$i(r)$	$j(r)$
1	3	2
2	3	5
3	4	5
4	3	4
5	2	5

La primera columna es el índice de cada nodo, y la segunda y tercera los índices de las neuronas alimentadoras.

La columna r indica el número de nodo, y la $i(r)$ y $j(r)$ son las que alimentan al nodo r . La segunda y la tercera columnas son generadas al azar. Una vez que las neuronas evalúan su función NAND, actualizan su estado de manera sincrónica. La Tabla 2 muestra la evolución a lo largo de cinco unidades de tiempo de la máquina de la Figura 6, a partir de una condición inicial.

Tabla 2. Una sucesión de estados del sistema de la máquina de la Figura 6.

1	1	1	0	0	1	0
2	1	1	1	0	1	0
3	0	1	1	1	1	1
4	0	1	0	1	0	1
5	1	0	1	0	1	0

En lenguaje moderno, la máquina desorganizada de tipo A de Turing es un sistema dinámico: dada una condición inicial, el vector de estado evoluciona y genera una sucesión de estados del sistema que son las columnas de la Tabla 2.

Pero la máquina A de Turing no tiene demasiadas posibilidades dinámicas: tiende a un estado estacionario; o bien, dada la finitud del número de nodos, desemboca en un estado periódico. Turing se dio cuenta de estas restricciones y modificó el modelo para dar lugar a las máquinas de tipo B. Al circuito de una máquina de tipo A, como el de la parte izquierda de la Figura 7, se le denota mediante un nodo cuadrado, como se puede apreciar en la parte derecha de la misma figura.

Con estos nuevos nodos es posible armar circuitos como el de la Figura 8, que se llamarán “máquinas desorganizadas de tipo B”. La Figura 8 ilustra una de ellas. Los circuitos de la parte izquierda de la Figura 7 tienen más posibilidades dinámicas, pues son capaces de introducir un retraso de una unidad de tiempo en la salida del nodo cuadrado. Esta situación se asemeja más a las neuronas reales, pero también permite que los nodos de las máquinas del tipo B tengan modificadores en sus señales de salida (las entradas de otras neuronas), y de esa manera se puede entrenar una red para lograr algún efecto deseado. Turing le llama “interferencia en el papel” (en nuestro lenguaje sería “cambio en el *software*”) para distinguirlo de un circuito que fuera modificado quitando o poniendo nodos (“interferencia con el desatornillador”).

En su ensayo Turing afirma que con esta modificación su máquina B puede empezar a organizarse y, eventualmente, dar lugar a una máquina inteligente. En esta fase, Turing ya no prosigue con las propuestas específicas de cómo diseñar un circuito y de cómo “interferir” con él; más bien da por hecho que se puede

construir, y pasa a discutir las posibilidades de sus máquinas. Incluso llega a proponer que sus máquinas son un cerebro sin cuerpo, y que son capaces de traducir y aprender lenguajes, dominar el cifrado de mensajes, hacer matemática y jugar ajedrez, *bridge*, etcétera.

El resto del ensayo es la exploración de todas las posibilidades que se abren con este tipo de propuesta teórica, y una buena discusión de la analogía entre máquina y cerebro. Propone que una de sus máquinas es como un bebé recién nacido, y muestra su visión al comentar los factores que inciden en su educación, como los elementos genéticos y los sociales.

No puedo afirmar si Turing perdió interés por los modelos neuronales, o si sus obligaciones laborales lo empujaron hacia otros derroteros. Lo cierto es que su manuscrito quedó guardado y no fue publicado sino hasta después de su muerte (tan tarde como en 1992). Turing nunca proporcionó un algoritmo de entrenamiento de sus máquinas. Ahora se sabe que la analogía entre una máquina desorganizada y el bebé recién nacido es válida: un bebé nace ciego y va adquiriendo el sentido de la vista conforme las neuronas establecen nuevas conexiones entre ellas, en respuesta a los estímulos visuales del entorno. Un niño, como Kaspar Hauser, sin la debida exposición al mundo y sus señales, no puede desarrollar capacidades sensoriales completas.

Los autores modernos han encontrado pequeños errores en el ensayo de Turing, pero ya no importa: la propuesta se hizo, se perdió y ya no tiene cabida en nuestro mundo, pues otros trabajos la superaron.

● La fábrica de los sueños

En 1957 Frank Rosenblatt propuso el *perceptrón*, una neurona artificial capaz de reconocer patrones con

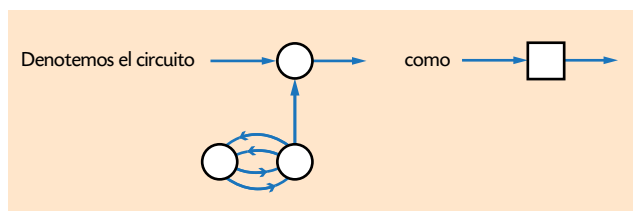


Figura 7. La redefinición de un nodo para incluir la posibilidad de “interferencia” o “reeducación”.

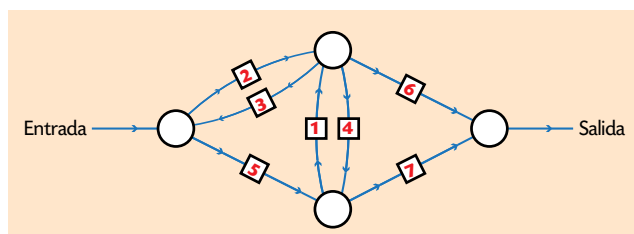


Figura 8. Una máquina desorganizada del tipo B.

separabilidad lineal. El trabajo de Rosenblatt se vio severamente inhibido por las críticas destructivas de Marvin Minsky, quien era el personaje más influyente en el naciente campo de la inteligencia artificial. Minsky, una fuerte personalidad con poder académico y político, bloqueó la investigación en el área del cómputo neuronal, y la retrasó al menos dos décadas. Los *mandarines de la ciencia*, como Minsky y Darwin nieta, han ejercido un poder político-académico perverso que obstaculizó el desarrollo del conocimiento.

En la segunda mitad de la década de 1980, una pléyade de investigadores se escapó de la trampa de Minsky y dieron el paso siguiente al generalizar el perceptrón y crear un *perceptrón multicapa*: un reconocedor general de patrones muy eficiente, con la posibilidad de manejar el caso de la separabilidad no lineal. A partir de esos años, y después de modelos muy exitosos (busque, por ejemplo, en la *Wikipedia*, “NetTalk”, una red de neuronas artificiales que aprende a hablar y que durante su fase de aprendizaje comete los mismos errores y balbuceos que los niños cuando adquieren el lenguaje), el desarrollo del cómputo neuronal ha sido vertiginoso, tanto en la teoría como en la práctica. Ahora tenemos identificadores de huellas digitales, reconocedores de música, clasificadores de imágenes, autos que se conducen solos, y decenas de ejemplos que ilustran el éxito de este campo.

Las consecuencias filosóficas no son desdeñables: una vez que sabemos que la separación mente-cuerpo es ficticia; ahora que todo el mundo sabe y acepta que nuestras funciones cognitivas, sensoriales y de comportamiento están mayormente ubicadas en el cerebro, las viejas preguntas regresan con mayor fuerza: ¿se llegará a entender el origen de la conciencia, de la imaginación, los sentimientos y los sueños? Todo parece indicar que la respuesta es positiva.

A partir del trabajo de Rosenblatt, a la línea de pensamiento que sugiere que estas interrogantes pueden responderse mediante el empleo de redes artificiales de neuronas se le llama *conexionismo*. No es la única escuela, ni sería deseable, pero al menos es una propuesta que eventualmente puede someterse a prueba y demostrar si sus planteamientos son correctos o no. De serlo, rasgos tan conocidos e inasibles como el libre albedrío, el amor y los sueños llegarán a comprenderse, y

tendremos un mundo mejor, dado que el conocimiento es bueno.

El gran trabajo visionario de Turing y las enormes contribuciones de los que lo han seguido han señalado la posibilidad de que, eventualmente, las máquinas inteligentes no se encuentren lejos. Quizá lo sepamos una vez que estén aquí (en 1964, Arthur C. Clarke escribió un relato breve: “Dial an F for Frankenstein”. En él, en un futuro distópico, la sociedad descubre que la red telefónica mundial ha sobrepasado una densidad crítica de interconexiones y una inteligencia suprahumana nace a partir de ese momento. Buena lectura).

Colofón

Nunca se construyó una máquina inteligente de Turing. Parte de la responsabilidad se le puede achacar a Charles Darwin nieta, pero el peso de la culpa recae en la sociedad inglesa que, como todos sabemos, orilló a Turing a una muerte prematura.

Pedro Miramontes es profesor de carrera del Departamento de Matemáticas de la Facultad de Ciencias de la Universidad Nacional Autónoma de México (UNAM), e integrante del Grupo de Biomatemáticas del mismo. Es investigador afiliado al Centro de Ciencias de la Complejidad (C3) de la UNAM.

pmv@ciencias.unam.mx

Lecturas recomendadas

Copeland, J. y D. Proudfoot (1999), “Alan Turing’s forgotten ideas in computer science”, *Scientific American*, 280(4):99-103.

McCullough, W. S. y W. H. Pitts (1943), “A logical calculus of the ideas immanent in nervous activity”, *Bulletin of Mathematical Biophysics*, 5:115-133.

Rosenblatt, F. (1961), *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*, Washington, Spartan Books.

Turing, A. M., “Intelligent machinery”, en Ince D. C. (editor, 1992), *Collected works of A. M. Turing - Mechanical Intelligence*, Elsevier.

