Cerebros, máquinas y mentes

Rosa María Mayorga *
Miami Dade College
rmayorga@mdc.edu

En los archivos digitales de la Open University[1], hay un vídeo de 1973 en el que se recoge un debate filosófico entre Gilbert Ryle, profesor de la Universidad de Oxford, y una joven profesora de la Universidad de Warwick sobre si el comportamiento inteligente puede explicarse sin hacer referencia a lo mental. Acosado por las preguntas incisivas y los comentarios perspicaces de su joven homólogo, Ryle admite que su explicación conductista/reduccionista, tal y como la propone en su libro *The Concept of Mind,* podría necesitar algunos ajustes. Al ver ese vídeo ahora, Susan Haack desearía haber «presionado más a Ryle»

> … sobre las «prácticas» o «convenciones» a las que apela como forma de completar su idea del comportamiento; y sobre el lugar que ocupa la neurofisiología. Hace tiempo que sospecho que tener una melodía en la cabeza o visualizar el rostro de un amigo es algo parecido a tener activadas aquellas neuronas que se dispararían si realmente se estuviera escuchando la melodía o viendo su rostro. Y no fui lo suficientemente dura con Ryle cuando se trató de la cuestión de si una máquina podía firmar su testamento: «Una máquina podría hacer una marca como el nombre que se le ha dado a un documento titulado "Última voluntad y testamento de A. Roe Bot"», debería haber dicho, «pero eso no sería firmar su testamento, independientemente de las convenciones legales que estableciéramos, porque una máquina no podría tener las intenciones, creencias, deseos, etc. adecuados… Quizás entonces podríamos haber avanzado hacia una mejor comprensión del fenómeno enormemente complejo de la mentalidad humana…».[2]

*Corresponding author.
[1] https://www.open.ac.uk/library/digital-archive/program/video:00525_3017
[2] https://www.academia.edu/81027914/That_Ryle_Interview_in_Retrospect

La cuestión de qué constituye la inteligencia y el comportamiento inteligente es especialmente apremiante hoy en día, y el interés va mucho más allá de los círculos filosóficos. Los avances en inteligencia artificial (IA) se producen más rápido de lo que podemos explicarlos. Estas formidables máquinas destacan en la realización eficaz de tareas específicas (inteligencia artificial estrecha o ANI), como diagnosticar cánceres, pilotar coches, componer música, dirigir armas, etc., que hasta hace poco solo podían realizar los seres humanos.[3] La llegada del Chat GPT de Open AI en 2022 y la rápida proliferación y mejora de estos grandes modelos de lenguaje (LLM) que pueden conversar con los seres humanos, tienen acceso a todo el conocimiento humano registrado y pueden dar respuesta a prácticamente cualquier pregunta que se les plantee, ha aumentado la urgencia de determinar qué es la inteligencia y el comportamiento inteligente. Las opiniones difieren sobre la fecha de llegada, pero los líderes de la industria y los investigadores afirman con confianza que la Inteligencia Artificial General (AGI), que iguala la inteligencia humana, es inminente, y que la Superinteligencia Artificial (ASI), que supera ampliamente la inteligencia de los más brillantes y dotados, le seguirá pronto, lo que hace saltar las alarmas, para algunos, de una amenaza existencial para la vida humana por parte de máquinas rebeldes.[4] Otros afirman que la AGI y la ASI no se producirán de forma independiente, ya que las máquinas siempre necesitarán la intervención humana para alcanzar el equivalente a la inteligencia humana.[5] ¿Qué criterios debería cumplir una máquina para tener un comportamiento inteligente? Para explorar esta cuestión, seguiré las pistas iniciales de Haack en su conversación con Ryle, concretamente, la relevancia de la neurofisiología en una explicación sobre la mente, con el fin de abordar la inteligencia artificial. La obra de Charles Peirce, fuente frecuente de inspiración para muchos, incluyendo a Haack, también tendrá un papel importante en este debate.

Cerebros

En *A Thousand Brains: A New Theory of Intelligence (Mil cerebros: una* nueva teoría *de la inteligencia)*, Jeff Hawkins, neurocientífico e ingeniero, propone una nueva forma de entender el oscuro proceso de cómo surge la inteligencia a partir de las células de nuestro cerebro; su objetivo final es aplicar este conocimiento para construir máquinas basadas en los mismos principios de la biología del cerebro (Hawkins 2021, 3). Su «nueva teoría de la inteligencia» ofrece interesantes perspectivas sobre nuestros sentidos de la visión, el tacto, el oído, el lenguaje y

[3] Y tareas que no realizan los seres humanos, como traducir la comunicación animal y desarrollar la comunicación entre especies; véase, por ejemplo, https://www.projectceti.org para conocer el trabajo sobre la aplicación del aprendizaje automático avanzado y la robótica de última generación para escuchar y traducir la comunicación de los cachalotes.

[4] Una figura destacada y muy vocal es Geoffrey Hinton, premio Nobel y «padrino de la IA», que aparece con frecuencia en los medios de comunicación: https://www.youtube.com/watch?v=IkdziSLYzHw

[5] Esta es una versión de la afirmación de Hawkins, tratada aquí.

el pensamiento abstracto, atributos que asociamos con la inteligencia y que son relevantes para nuestro debate.

Hawkins comienza con algunos conceptos básicos de biología evolutiva. Al igual que el cerebro de la mayoría de los mamíferos complejos, el cerebro humano desarrolló la capacidad de realizar comportamientos cada vez más sofisticados añadiendo, en lugar de sustituyendo, nuevas partes a las partes más antiguas que controlan funciones básicas fundamentales para la supervivencia, como la respiración, la alimentación, la procreación, los reflejos, etc. La parte más nueva de nuestro cerebro, el neocórtex, es la capa externa que envuelve el tronco cerebral, el tálamo y la amígdala, más antiguos; el neocórtex ocupa alrededor del 70 % del volumen, lo que, si se extendiera en plano, sería de unos 24 centímetros cuadrados y unos 2,5 milímetros de grosor (Hawkins 2021, 6). El neocórtex es el órgano de la inteligencia, donde residen las capacidades para las funciones cognitivas como la visión, el lenguaje, el razonamiento, etc. El cerebro «nuevo» y el «antiguo» están conectados a través de fibras nerviosas para funcionar de forma cooperativa como uno solo, con el neocórtex enviando señales eléctricas como órdenes al cerebro antiguo, que a su vez controla todos los músculos, como los que intervienen en la respiración, que no requieren ninguna entrada del neocórtex. A diferencia del cerebro antiguo, con áreas visualmente diferenciadas responsables de funciones corporales primitivas (automáticas) como la digestión, así como de reacciones reflejas y emocionales asociadas a la supervivencia, como la agresividad, el neocórtex es sorprendentemente diferente, sin divisiones visualmente evidentes en los pliegues grises ondulados, aunque hay diferentes regiones responsables de las diferentes funciones de la visión, el tacto, el lenguaje, la reflexión, etc. (Hawkins 2021, 12-14).

Esta homogeneidad inesperada en el aspecto exterior general del neocórtex se repite también en el interior. La primera persona que observó con detalle al microscopio las células cerebrales individuales con una técnica de tinción especial fue el médico y científico Santiago Ramón y Cajal, a finales del siglo XIX (Hawkins 2021, 12-14). Ramón y Cajal utilizó estas tinciones para crear imágenes complejas de cada parte del cerebro, que mostraron por primera vez que las células cerebrales, al igual que otras células, tenían un núcleo, una membrana celular que definía su límite, que se extendía con apéndices ramificados llamados axones y dendritas. Las ramificaciones dendríticas se agrupan cerca de la célula y el axón establece muchas conexiones con las neuronas cercanas, pero también con las más distantes, viajando desde el neocórtex hasta la médula espinal. Hoy en día sabemos, gracias a las imágenes cerebrales, que las neuronas crean picos, o señales eléctricas que comienzan cerca del cuerpo celular y viajan a lo largo del axón hasta llegar al final de cada ramificación. Las neuronas establecen conexiones (sinapsis) con las dendritas de otras neuronas; cuando un pico que viaja a lo largo de un axón llega a una sinapsis, libera una sustancia química que entra en la dendrita de una neurona receptora. (Hawkins 2021, 44)

Hawkins propone que el neocórtex creció hasta constituir el 70 % de nuestro cerebro, no creando nada nuevo desde el punto de vista evolutivo, sino copiando el mismo circuito una y otra vez. Aunque nuestros neocórtex son más grandes que los de un perro o una rata, todos están formados por el mismo elemento; simplemente

tenemos más copias de células, y todas las partes del órgano funcionan según el mismo principio. Fundamentalmente, esto significa que todas las cosas que consideramos relacionadas con la inteligencia, como la percepción, el lenguaje y el pensamiento de alto nivel, utilizan el mismo proceso. La razón por la que todas las regiones parecen iguales es que todas hacen lo mismo; lo que las diferencia no es su función en sí, propone Hawkins, sino más bien a qué están conectadas. La región conectada a los ojos produce la visión; la conectada a los oídos produce la audición; y así sucesivamente. Al igual que Darwin propuso un mecanismo (algoritmo), es decir, la supervivencia del más apto para explicar la extraordinaria diversidad de la vida, Hawkins propone que todas las cosas asociadas con la inteligencia son en realidad «el mismo algoritmo cortical subyacente‘ (Hawkins 2021, 24). Y la ubicación de este algoritmo cortical, esta unidad de inteligencia, es la «columna cortical», que se extiende a lo largo de todo el grosor del neocórtex, una de las aproximadamente 150 000 columnas apiladas verticalmente una al lado de la otra, como cortos fideos espagueti. Estas columnas no son visibles al microscopio, ni tienen límites visibles; su existencia se postula porque, como se observa en las imágenes cerebrales, las neuronas de una

columna, o hebra, responden con impulsos eléctricos (picos) a los estímulos procedentes de un área específica de nuestros órganos sensoriales. Hawkins sostiene que su afirmación de que el cerebro se basa en un método polivalente para el aprendizaje también explica nuestra flexibilidad para aprender cosas nuevas no relacionadas entre sí, desde hacer un soufflé hasta viajar a la luna (Hawkins 2021, 25-27).

Pero, ¿por qué hay tantas columnas corticales? ¿Y para qué sirven? La respuesta, según Hawkins, tiene que ver con el valor de la predicción para mejorar la supervivencia: por ejemplo, si podemos adivinar con precisión de dónde proviene el sonido de un choque, podríamos actuar en consecuencia, es decir, agacharnos a tiempo. Para hacer predicciones, el cerebro tiene que aprender el *statu quo*, es decir, lo que se espera, basándose en experiencias previas (Hawkins 2021, 31). Y una forma eficaz de hacerlo es crear un modelo del mundo y basar las predicciones en este modelo. El cerebro crea este modelo prácticamente desde cero desde el nacimiento: el marco básico está ahí, pero construimos el contenido a través de la experiencia sensorial, por ejemplo, qué es un árbol, los diferentes patrones y colores de las hojas entre un roble y una palmera, el sonido que hacen las bellotas al caer sobre el techo en comparación con los cocos. El número de cosas que hay en el modelo es enorme, ya que representa absolutamente todo lo que hemos aprendido sobre el mundo; pero eso no es lo único que hace que el modelo sea increíblemente intrincado y complejo: el mundo está en constante cambio, por lo que el modelo debe actualizarse constantemente para poder hacer predicciones acertadas. Por lo tanto, se necesitan muchas columnas corticales.

Nuestra información sensorial cambia constantemente debido al movimiento: el movimiento de las cosas en el mundo (los árboles se mecen con el viento, el lápiz rueda por el escritorio, el día se convierte en noche), pero también porque nos movemos por el mundo: cada vez que damos un paso, cambiamos la mirada, movemos la cabeza o cambiamos de dirección, experimentamos una nueva sensación desde un ángulo diferente. Y con cada movimiento, el cerebro

predice cuál será la siguiente sensación. Mientras camino hacia la puerta de mi casa, espero ver el alto roble a mi izquierda y las palmeras a mi derecha; no espero ver varios objetos redondos grandes en mi camino que me bloqueen el paso. Mi modelo necesita actualizarse: evito los cocos caídos y tomo nota mentalmente de llamar al servicio de poda de árboles. La teoría de Hawkins es que el cerebro realiza cientos, si no miles, de predicciones en un momento dado, pero solo nos damos cuenta cuando hay un error en la predicción. Concluye que hay dos tipos de predicciones: las que se refieren a cambios en los objetos del mundo y las que se refieren a cambios resultantes de nuestro movimiento en el mundo (Hawkins 2021, 41). Para hacer predicciones sobre nuestros movimientos, es esencial una descripción precisa de nuestra ubicación en el espacio, no solo del cuerpo entero, sino también de partes específicas en relación con los objetos, por ejemplo, mi mano derecha cuando busco la llave de mi casa en el bolso. ¿Cómo puede mi cerebro predecir que mi mano va a coger las llaves del bolsillo del bolso donde siempre las guardo? Necesitaría saber dos cosas: qué objeto está tocando y la ubicación de mi mano en relación con el objeto.

Hawkins postula que la forma en que las neuronas representan esta información es a través de un marco de referencia (como una cuadrícula en un mapa) del objeto, en este caso, la llave, y la ubicación de los dedos de mi mano mientras agarran la llave en el bolso. Las neuronas,

entonces, procesan la información sensorial creando marcos de referencia para los objetos y rastreando la ubicación del sensor; una sola columna cortical aprende la forma tridimensional de los objetos al detectar y rastrear el movimiento constantemente. Hawkins postula que la mayoría de las predicciones comienzan cuando un conjunto de sinapsis cercanas entre sí en una rama dendrítica recibe información al mismo tiempo, reconociendo un patrón de actividad en otras neuronas, lo que provoca un pico dendrítico. El pico viaja al cuerpo celular, poniéndolo en un estado predictivo al elevar temporalmente el voltaje y preparándolo para dispararse antes que las otras neuronas (Hawkins 2021, 46). Si llega la información prevista (el objeto en el bolsillo de mi bolso se siente como la llave de la casa), inhibe el disparo de otras neuronas. Las predicciones no se envían a través del axón a otras neuronas, lo que explica por qué no somos conscientes de la gran mayoría de la actividad neuronal. Sin embargo, si llega una entrada inesperada (un objeto largo y puntiagudo, en lugar de uno más corto y romo), varias neuronas se activan a la vez, lo que provoca mucha más actividad eléctrica y muscular, llamando mi atención y llevándome finalmente a la conclusión de que el objeto que hay en el bolsillo del bolso no es mi llave, sino mi lápiz de sombra de ojos, que había perdido.

Navegar con éxito por el mundo es esencial para la supervivencia de todas las criaturas, por lo que, desde muy temprano, los mamíferos desarrollaron estas neuronas creadoras de mapas que proporcionan este tipo de sistema básico de navegación interna. Nosotros también tenemos células de red y células de lugar en el hipocampo y la corteza entorrinal, parte de nuestro cerebro antiguo; en su mayoría rastrean el mecanismo de supervivencia más rudimentario, es decir, la ubicación del cuerpo en el entorno (Hawkins 2021, 61). El neocórtex, con básicamente las mismas células, tiene una estructura diferente, con 150 000 copias

de este mismo circuito, una por cada columna cortical, por lo que rastrea miles de ubicaciones de nuestro cuerpo simultáneamente. Cada pequeña zona de la piel, cada parte de la retina, tiene su propio marco de referencia y también modela objetos completos (Hawkins 2021, 62-63).

Si Hawkins está en lo cierto y cada columna cortical realiza la misma función básica, entonces el lenguaje y otras capacidades cognitivas de nivel superior que asociamos con la inteligencia son, en algún nivel fundamental, similares a la percepción. Al igual que los marcos de referencia organizan la información que percibimos, Hawkins propone que los marcos de referencia organizan la información que no percibimos directamente, por ejemplo, los átomos o los genes, tratando de visualizarlos. Razonamos sobre conceptos abstractos como la belleza, la justicia o la libertad, por lo que los marcos de referencia también deben estar involucrados si existe un algoritmo común para todas las columnas corticales. Una explicación evolutiva sería algo así: los cerebros desarrollaron primero marcos de referencia para aprender la estructura de los entornos, de modo que pudiéramos movernos eficazmente por el mundo. Luego, nuestros cerebros evolucionaron para utilizar el mismo mecanismo para aprender la estructura de los objetos físicos, de modo que pudiéramos reconocerlos y manipularlos. Y luego los cerebros evolucionaron de nuevo para adaptar el mismo mecanismo para aprender y representar la estructura de los objetos conceptuales subyacentes (Hawkins 2021, 69-71).

El pensamiento, entonces, es una forma de movimiento. Si todo el conocimiento se almacena en ubicaciones relativas a marcos de referencia, entonces, para acceder al conocimiento almacenado, tenemos que activar la ubicación adecuada en el marco de referencia pertinente.

El pensamiento se produce cuando las neuronas activan una ubicación tras otra en un marco de referencia, recuperando lo que se almacenó en cada ubicación. La sucesión de pensamientos se corresponde con la sucesión de sensaciones al tocar un objeto con los dedos, o la sucesión de imágenes que vemos al entrar en una habitación, o la sucesión de imágenes de objetos conceptuales. Si esto es así, entonces lo que llamamos pensar es en realidad moverse a través del espacio (Hawkins 2021, 80). De manera similar, Hawkins explica la conciencia, el sentido de la percepción formado por los recuerdos momento a momento de las acciones; esto, explica, se activa mediante neuronas secuenciales que representan experiencias pasadas y presentes. Ahora bien, dominar el conocimiento conceptual, o alcanzar la conciencia, es decir, organizar los marcos de referencia adecuados, no es tan fácil como organizar los marcos de referencia para los objetos físicos; es probable que la razón sea que esta etapa es la más reciente en el proceso evolutivo y ha habido menos tiempo para su desarrollo.

Máquinas

La última ola de nuevas tecnologías que prometía ponernos en el camino hacia la creación de máquinas inteligentes son las «redes neuronales artificiales», a menudo denominadas «aprendizaje profundo». Este nuevo enfoque para codificar el conocimiento entrena a las máquinas con gran cantidad de texto, palabra por

palabra, agotando todo el contenido del Internet. Al entrenarse de esta manera, las redes aprenden la probabilidad de que ciertas palabras sigan a otras, basándose en estadísticas y en una inmensa cantidad de datos, y se vuelven cada vez más competentes para responder a cualquier pregunta que se les plantee[6] . Pero a pesar de los resultados increíblemente impresionantes de Chat GPT y otras máquinas que utilizan métodos de aprendizaje profundo, como los coches autónomos, los ordenadores que juegan al ajedrez, etc., Hawkins no cree que las máquinas actuales sean «verdaderamente inteligentes», ni que posean conocimiento; en su opinión, poseer verdadero conocimiento es algo más que tener listas de datos. Más bien, el conocimiento requiere representar los hechos cotidianos «de una manera útil», y hasta ahora los investigadores de IA no han sido capaces de resolver este «problema de representación del conocimiento» (Hawkins 2021, 122). Hawkins cree que el objetivo de la IGA no se alcanzará a menos que las máquinas modelen el mundo de la misma manera que lo hace el cerebro, concretamente el neocórtex, utilizando marcos de referencia similares a mapas que puedan manipularse virtualmente; sin embargo, está convencido de que esto se puede lograr en un futuro no muy lejano y que las máquinas inteligentes también serán conscientes.[7]

Dado que define la inteligencia como «la capacidad de un sistema para aprender un modelo del mundo» y que el neocórtex, el «órgano de la inteligencia», es el principal responsable de crear este modelo, Hawkins sostiene que, al construir máquinas inteligentes, lo que hay que replicar es el equivalente al nuevo cerebro, o neocórtex, y no al cerebro antiguo. Compuesto por diversas partes, el cerebro antiguo regula las funciones vitales básicas y el comportamiento, incluyendo la respiración, la digestión, los latidos del corazón y otros movimientos automáticos, así como los instintos, las emociones y las motivaciones asociadas con la supervivencia y la procreación, y Hawkins sostiene que sería demasiado complejo, demasiado difícil y probablemente peligroso de replicar. Aunque mucho más grande en superficie, el neocórtex es mucho más simple en su estructura, compuesto básicamente por muchas copias de un elemento, la columna cortical, y por lo tanto mucho más fácil de recrear en una máquina. Una máquina inteligente necesitará entonces tener sensores conectados al movimiento y un mapa, o modelo, de su entorno. Podría tener diferentes formas de percibir su entorno, navegar y llevar a cabo tareas mecánicas, desde robots mecanizados a escala humana hasta nanorobots híbridos que administran tratamientos médicos a nivel micro a través de «células» controladas.[8] Ahora bien, dado que el antiguo complejo cerebral es la sede de las motivaciones, los deseos, las emociones, los instintos, etc.,

[6]No siempre son precisas o veraces, y «alucinan» o dan respuestas incorrectas.

[7]Algunas máquinas, como los ordenadores que juegan al ajedrez, los coches autónomos y los robots, ya utilizan marcos de referencia de forma limitada: el tablero de ajedrez es un marco de referencia que representa la ubicación de cada pieza y sigue sus movimientos; los coches autónomos utilizan el GPS, el sistema de seguimiento por satélite que muestra la ubicación del coche, como marco de referencia; los robots están equipados con marcos de referencia para poder desplazarse de un lugar a otro y/o mover piezas en relación con el entorno. Sin embargo, todas estas máquinas son ejemplos de sistemas «dedicados», que son específicos para una

[8]Para ver un ejemplo reciente de robots microscópicos, consulte https://www.nbcnews.com/mach/science/these-tiny-robots-could-be-disease-fighting-machines-inside-body-ncna861451

impulsados principalmente por la supervivencia, y que influyen en gran medida en el comportamiento, los mecanismos para los objetivos y las motivaciones, así como las medidas de seguridad,[9] deberán ser diseñados, programados y supervisados en la máquina directamente por los seres humanos, ya que estos no surgirán por sí solos en las máquinas creadas según el modelo neocortical propuesto por Hawkins.[10]

tarea, y no sistemas «universales» o «de uso general» que pueden aprender, como nosotros, muchas cosas sin borrar la memoria y empezar de nuevo (Hawkins, 2021, 127).

Peirce

Charles Peirce (1839-1914) era filósofo y científico de formación, por lo que estaba convencido de primera mano de la importancia de que la teoría científica y la teoría filosófica se influyeran mutuamente. Se mantenía al día de los últimos trabajos de sus colegas en filosofía y en ciencias, como se evidencia en sus numerosas reseñas de libros y publicaciones, así como en sus contribuciones al Century Dictionary. Como tal, abordó el tema de las «máquinas lógicas» en el contexto de algunos intentos actuales de mecanización de tareas sencillas, en un artículo para la revista American Journal of Psychology en 1887 (CN 1:41). Aunque lo que llamamos IA no surgió hasta varias décadas después, sus comentarios son pertinentes para nuestro debate aquí, es decir, qué constituye la inteligencia y el comportamiento inteligente (humano o mecánico) y la relevancia de la neurofisiología en nuestra investigación.[11] Comenzaré por lo segundo.

Peirce conocía el trabajo de Ramón y Cajal sobre las neuronas; no está claro cuándo lo descubrió, pero asistió a las conferencias de Ramón y Cajal en la Universidad de Clark en 1899 y escribió un artículo para *Science* sobre el volumen conmemorativo de las conferencias.[12] Sin embargo, es evidente por sus escritos anteriores que Peirce conocía la neurofisiología básica; sus comentarios anticipan y

[9] Hawkins sugiere las tres leyes de la robótica de Isaac Asimov: un robot no puede dañar a un ser humano; un robot debe obedecer las órdenes de un ser humano; un robot debe proteger su propia existencia siempre que no entre en conflicto con la primera o la segunda regla (Hawkins, 2021, 154).

[10] Hawkins considera innecesario, inviable y quizás peligroso intentar recrear e incorporar emociones en las máquinas. Esto recuerda a la «paradoja de Moravec», que surge porque el razonamiento de alto nivel (por ejemplo, jugar al ajedrez, demostraciones matemáticas), que los humanos consideran «complejo», resulta ser relativamente más fácil de codificar en los sistemas de IA (lógica, inferencia), mientras que las habilidades de bajo nivel (por ejemplo, correr, reconocer rostros), que los humanos consideran «simples», resultan extraordinariamente difíciles de replicar en las máquinas.https://medium.com/intuitionmachine/moravecs-paradox-from-the-4e-cognitive-psychology-view-539359b432fa

[11] La teoría semiótica de Peirce, la lógica de las relaciones, los gráficos existenciales, las teorías de la inferencia, así como sus comentarios sobre las «máquinas de razonar» han dado lugar a una importante bibliografía teórica y le han valido el reconocimiento como precursor de la IA moderna.

[12] MS 298, p. 11bis, c.1906; https://www.unav.es/gep/ClarkUniversity89-99.html](https://www.unav.es/gep/ClarkUniversity89-99.html)

se asemejan a algunas de las observaciones de Hawkins sobre el cerebro antiguo y el cerebro nuevo, la función del sistema nervioso, la similitud entre la percepción y la formación de conceptos, y el mecanismo de predicción.

Aunque Peirce no hace la distinción entre cerebro antiguo y nuevo *per se*, sí distingue entre las dos funciones. Los procesos corporales reflexivos o automáticos, tales como «la respiración, la circulación y la digestión, se llevan a cabo tal cual, sin ninguna interferencia de la razón… y se realizan de forma menos deficiente de forma inconsciente que si estuvieran sometidos al régimen de la lógica;» (CP 7.448) y «se forman un número inmenso de asociaciones, que permanecen mientras duran, en el fondo de la conciencia, es decir, en la oscuridad subjetiva. Pero tan pronto como se produce una sugerencia cerebro-motora… la idea de ejercer voluntariamente el pensamiento» (lo que ocurriría en el neocórtex, según Hawkins), «todo el conjunto se ilumina» (CP 7.434).

Según Peirce, existen «tres funciones fundamentales del sistema nervioso, a saber: en primer lugar, la excitación de las células; en segundo lugar, la transferencia de la excitación a través de las fibras; en tercer lugar, la fijación de tendencias definidas bajo la influencia del hábito… pero todas ellas pueden resumirse bajo los conceptos de sensibilidad, movimiento y crecimiento» (CP 1.393).

La actividad sináptica, descrita como «una excitación de los nervios», se propaga, «afectando cada vez más a la materia nerviosa», pero «no de manera uniforme… sino de forma muy diferente en distintas direcciones», y la dirección exacta que tomará la propagación parece tan incierta como el lanzamiento de unos dados» (CP 6.280). La «excitación total puede aumentar considerablemente» y propagarse, o «también puede agotarse» (CP 7.396). Peirce también es consciente de que los nervios controlan diferentes procesos en función de sus conexiones (la noción de Hawkins del cerebro antiguo o nuevo) «desde los nervios espinales, ya sea «en contracciones de músculos voluntarios, acompañadas en su mayoría de una actividad mental sobria», o «en secreciones glandulares, acciones sobre músculos involuntarios, probablemente a través de los nervios simpáticos, y estos casos suelen ir acompañados de una excitación emocional del alma» (CP 7.396).

En términos de sensibilidad o sensaciones, Peirce intenta explicar el mecanismo de la percepción visual y cómo vemos una imagen continua: «la retina está formada por innumerables agujas que apuntan hacia la luz… supongamos que cada uno de esos puntos nerviosos transmite la sensación de una pequeña superficie cubierta… lo que debemos ver inmediatamente debe ser… entonces… una colección de puntos… pero vemos una superficie continua… por lo tanto, la suma de estas impresiones es una condición necesaria de cualquier percepción producida por la excitación… de cada una [de las células nerviosas]» (CP 5.223). Al igual que Hawkins, Peirce compara el proceso de formación de conceptos con el de la percepción, explicando que «dos ideas son similares en la medida en que las mismas células nerviosas han participado en su producción… esa identidad radica en que una parte de una idea y una parte de la otra idea son la sensación peculiar de la excitación de una o más células nerviosas» (CP 1.388). Así es como crecen las conexiones nerviosas.

Una característica especial de los nervios, según Peirce, es su capacidad para estar «especialmente preparados para adoptar y cambiar sus hábitos» (CP 6.281). Ahora bien, «si la misma célula que una vez se excitó… a lo largo de una deter-

minada vía o vías… se excita por segunda vez, es más probable que se descargue por segunda vez a lo largo de algunas o todas esas [mismas] vías que si no se hubiera descargado antes» (CP 1.390). Hawkins explicaría esta repetición de un pico en la misma vía la próxima vez que se presentara un estímulo similar como las neuronas que predicen la siguiente acción. Peirce lo explica diciendo que las células nerviosas adquieren hábitos y actúan de esta manera para resolver, y por lo tanto eliminar, la irritación provocada por el estímulo (CP 6.281).[13] Peirce supone que la eliminación de la irritación, especialmente cuando está relacionada con la nutrición en los animales, a su vez provoca placer. «De hecho», dice Peirce, «la mayor parte de las acciones inteligentes están dirigidas a provocar el cese de alguna irritación. Comemos para librarnos del hambre, etc.» (CP 6.282). Y luego añade, de forma bastante inesperada: «Parecería posible que la tendencia a actuar de forma inteligente, es decir, con el fin de obtener un resultado determinado, pudiera surgir en un sistema meramente mecánico» (CP 6.285). Aunque esto podría ser un «absurdo monstruoso», sin embargo, se siente «bastante seguro de que la hipótesis ofrece un punto de vista instructivo desde el que contemplar la cuestión general… de que la inteligencia… consiste… en actuar de una determinada manera (CP 6.286). Concretamente, Peirce afirma en otra parte que la inteligencia es «capaz de aprender por experiencia» y se refiere a ella como «inteligencia científica» (CP 2.227).

Mientras que los «sistemas mecánicos» que Peirce menciona anteriormente son orgánicos, o están hechos de «protoplasma», desde células simples hasta criaturas vivientes más complejas, como las ranas, en su artículo de 1887 Peirce aborda las «máquinas lógicas». Comienza afirmando que «la cuestión de cuánto del trabajo de pensar podría hacer una máquina y qué parte debe dejarse para la mente viva no carece de importancia práctica concebible» y, además, «su estudio puede… arrojar la luz necesaria sobre la naturaleza del proceso de razonamiento» (Peirce 1887, 165). Aquí vemos un contraste entre Hawkins, el neurocientífico, y Peirce, el filósofo; mientras que Hawkins estudia el cerebro con el fin de crear máquinas inteligentes, Peirce ve el desarrollo de máquinas lógicas como una oportunidad para explorar más a fondo la naturaleza del razonamiento y la inteligencia.[14]

Peirce continúa diciendo que no es absurdo suponer que un «instrumento» puede hacer el trabajo de una mente y analiza las máquinas mecánicas inventadas por William Stanley Jevons (el «piano lógico») y posteriormente simplificadas por Allan Marquand, que, mediante un teclado, palancas internas y poleas, «reciben las premisas y dan las conclusiones mediante el giro de una manivela» (Peirce 1887, 165). Considera que estas, así como las «máquinas matemáticas» o calculadoras

[13] Véase también el paralelismo con la creencia y la duda (CP 5.373).

[14] La filósofa y científica cognitiva británica Margaret Boden, recientemente fallecida, no era experta en el uso de ordenadores, pero «utilizó el lenguaje de los ordenadores para explorar la naturaleza del pensamiento y la creatividad, lo que la llevó a obtener conocimientos proféticos sobre las posibilidades y limitaciones de la inteligencia artificial… Produjo varios libros, entre los que destacan *The Creative Mind: Myths and Mechanisms* (1990) y *Mind as Machine: A History of Cognitive Science* (2006)— que contribuyeron a dar forma al debate filosófico sobre la inteligencia humana y artificial durante décadas». https://www.nytimes.com/2025/08/14/science/margaret-boden-dead.html

de Webb y Babbage, realizan «razonamientos que no son sencillos», y afirma que el «secreto de todas las máquinas de razonamiento» es muy simple: «el mismo principio que subyace a toda álgebra lógica… que cualquier relación entre los objetos sobre los que se razona está destinada a ser el eje de un razonamiento, esa misma relación general debe poder introducirse entre ciertas partes de la máquina…» (Peirce 1887, 167-68). Lo que Peirce quiere decir con esto es que, al igual que el álgebra utiliza símbolos para representar y manipular relaciones matemáticas, una máquina que pueda representar relaciones lógicas, o inferencias, como las que existen entre premisas y conclusiones, realizará un tipo de razonamiento. Por esta razón, denomina «lógicas» a estas máquinas computacionales, ya que proporcionan el equivalente a tablas de verdad lógicas para las afirmaciones y ecuaciones limitadas que se les introducen. Peirce compara esto con el razonamiento humano y cómo ambos muestran la capacidad de extraer inferencias —

> Cuando razonamos con nuestra mente sin ayuda, hacemos básicamente lo mismo, es decir, construimos una imagen en nuestra imaginación bajo ciertas condiciones
>
> generales y observamos el resultado. Desde este punto de vista, todas las máquinas son máquinas de razonar, en la medida en que existen ciertas relaciones entre sus partes, relaciones que implican otras relaciones que no se pretendían expresamente (Peirce 1887, 167-68).[15]

Las matemáticas y la lógica están interconectadas.[16] «La fuente principal de la verdad lógica», afirma Peirce, «aunque nunca reconocida por los lógicos, siempre ha sido y siempre debe ser la misma que la fuente de la verdad matemática… y la verdad matemática se deriva de la observación de las creaciones de nuestra propia imaginación visual, que podemos plasmar en papel en forma de diagramas». Estos diagramas, por supuesto, son los famosos gráficos existenciales que desarrolla Peirce. Hawkins reconoce la conexión con las matemáticas en su explicación de cómo funciona el cerebro en términos de movimiento:

> En matemáticas, tenemos equivalentes de movimientos. Son operadores que se realizan. «Voy a tomar esto y lo voy a manipular con una transformada de Fourier». «Voy a manipularlo con una división» o algo por el estilo. Y entonces acabas en un espacio diferente. Y así, el cerebro descubre un modelo estructural de las matemáticas. Si hablas con matemáticos, ellos construyen una estructura física de estructuras matemáticas. No es como si yo simplemente aplicara alguna fórmula. Ellos perciben intuitivamente la estructura de las matemáticas… Está muy claro que la mente utiliza el mismo mecanismo para todo: para las matemáticas, el lenguaje… y así sucesivamente.[17]

[15] Aquí vemos de nuevo la afirmación de que pensar es como percibir.

[16] Véase «Why Study Logic?» (¿Por qué estudiar lógica?), de Mark Migotti, en *The Oxford Handbook of Charles S. Peirce,* 2024.

[17] https://www.sfexaminer.com/news/technology/why-palm-founder-jeff-hawkins-next-big-thing-could-be-ai/article_b0aede7a-3021-11ef-a127-9f48a3ca5ee8.html

Para Peirce, la lógica, las matemáticas y la semiosis están todas conectadas: «La lógica, en su sentido general, es… solo otro nombre para la semiótica… la doctrina formal de los signos» (CP 2.227). Describe el proceso de abstracción, o formación de conceptos, como algo que también implica «una especie de observación… observamos los caracteres de los signos que conocemos y, a partir de esa observación… llegamos a afirmaciones… a lo que deben ser los caracteres de todos los signos utilizados por una inteligencia «científica», es decir, por una inteligencia capaz de aprender por experiencia» (CP 2.227). Da un ejemplo:

> Es una experiencia familiar para todo ser humano desear algo que está muy por encima de sus posibilidades actuales y, a continuación, preguntarse: «¿Desearía lo mismo si tuviera los medios suficientes para satisfacer ese deseo?». Para responder a esa pregunta, él… hace en su imaginación una especie de diagrama esquemático, o boceto, de sí mismo, considera qué modificaciones requeriría la situación hipotética en esa imagen y luego la examina, es decir, observa lo que ha imaginado para ver si se percibe el mismo deseo ardiente. Mediante ese proceso, que en el fondo se parece mucho al razonamiento matemático, podemos llegar a conclusiones sobre lo que sería cierto en todos los casos en relación con los signos, siempre que la inteligencia que los utilice sea científica (CP 2.227).[18]

En otra parte, Peirce repite la comparación entre el hombre y la máquina. Así como «el resultado que produce la máquina lógica tiene una relación con los datos con los que se le alimentó… de la misma manera, un hombre puede ser considerado como una máquina que produce una frase escrita que expresa una conclusión… habiéndosele alimentado con una declaración escrita de hechos, como premisa» (CP 2.59). Este rendimiento, insiste, «no tiene ninguna relación esencial con el hecho de que la máquina funcione con ruedas dentadas, mientras que el hombre funciona con una disposición poco conocida de células *cerebrales»*, y por lo tanto «no tiene nada que ver con la crítica lógica, que es igualmente aplicable al rendimiento de la máquina y al del hombre; es simplemente la cuestión de si la conclusión puede ser falsa mientras la premisa es verdadera» (CP 2.59). Peirce afirma que «aunque no todo razonamiento es cálculo, es ciertamente cierto que el cálculo numérico es razonamiento» . Todas esas máquinas «realizan inferencias; y esas inferencias están sujetas a las reglas de la lógica» (CP 2.56).

Ahora bien, el razonamiento, o inferencia, «es de tres tipos elementales… inducción, deducción y… abducción» (CP 2.774). Peirce considera que las máquinas lógicas de su época realizaban silogismos deductivos simples, tales como: «Si A, entonces B; si B, entonces C; por lo tanto, si A, entonces C», siempre y cuando creemos las conexiones adecuadas en las partes de la máquina que representan

[18] Consideremos también: «El pensamiento matemático avanza principalmente mediante la generalización; y las conclusiones generalizadas se hacen rigurosamente lógicas mediante el recurso de generalizar correspondientemente las premisas» (CN 2:85). Y: «Entre las inferencias menores y las mayores se encuentra una clase que se rige mejor por los hábitos, pero por hábitos formados o corregidos bajo la crítica consciente» (CP 7.450).

A, B y C (Peirce 1887, 167). Si bien se puede decir que estas máquinas realizan deducciones simples, no realizan inducciones ni abducciones.[19]

Aunque Peirce cree que las máquinas realizan un tipo de razonamiento, no cree que se comporten de forma inteligente, ya que eso implicaría aprender de la experiencia. Pero, ¿diría Peirce lo mismo sobre las impresionantes máquinas LLM actuales, como Chat GPT, con logros tan impresionantes? Veamos lo que dice Peirce sobre la inteligencia y el conocimiento.

El «proceso de pensamiento real… comienza en las percepciones mismas. Pero una percepción no puede representarse con palabras y, en consecuencia, la primera parte del… pensamiento no puede representarse mediante ninguna forma lógica de argumento. Nuestra explicación lógica del asunto tiene que partir de un hecho perceptivo, o de una proposición resultante del pensamiento sobre una percepción… El… argumento lógico solo representa la última parte del pensamiento… (CP 2.27). [Y] «la lógica debe comenzar con una crítica del conocimiento…» (CP 2.63).

> Nuestro conocimiento de cualquier tema nunca va más allá de recopilar observaciones y formar algunas expectativas semiconscientes, hasta que nos enfrentamos a alguna experiencia contraria a esas expectativas. Eso nos despierta de inmediato a la conciencia… hacemos una suposición, una abducción, la única forma de inferencia que realmente puede introducir nuevas ideas y generar nueva información (CP 7.36).

Peirce sostiene que este instinto de conjetura, o «*il lume naturale*», es un rasgo evolutivo que nuestra mente ha desarrollado como resultado de nuestra experiencia como parte del universo,[20] y bajo la influencia de estos mismos principios, nos inclina a aceptar una hipótesis como verdadera o probable. Una vez que hemos formado una explicación probable, pasamos a la evidencia física y vemos si la hipótesis se verifica o no mediante la observación en el mundo real, que es el proceso de inducción. Así es como nos acercamos a la verdad.[21]

---

[19] «Los tipos superiores de razonamiento relativos a términos relativos no pueden (por lo que podemos ver hasta ahora) realizarse mecánicamente» (CN 1:87).

[20] Consideremos: «El tercer tipo de razonamiento prueba lo que puede hacer *il lume naturale*, que iluminó los pasos de Galileo. En realidad, es una apelación al instinto. Así, la razón, a pesar de todos los adornos que suele llevar, en las crisis vitales recurre a su médula para implorar el socorro del instinto» (CP 1.630). Y: «En la evolución de la ciencia, la conjetura desempeña el mismo papel que las variaciones en la reproducción en la evolución de las formas biológicas, según la teoría darwiniana» (CP 1.38).

[21] «… hay que reconocer que si supiéramos que el impulso de preferir una hipótesis a otra fuera realmente análogo a los instintos de las aves y las avispas, sería una tontería no darle rienda suelta, dentro de los límites de la razón; sobre todo porque debemos plantearnos alguna hipótesis, o de lo contrario renunciar a todo conocimiento más allá del que ya hemos adquirido por ese mismo medio. Pero, ¿es cierto que el hombre posee esta facultad mágica? No, respondo, en la medida en que acierta a la primera, ni quizá a la segunda; pero que una mente bien preparada ha adivinado maravillosamente pronto cada secreto de la naturaleza es una verdad histórica. Todas las teorías de la ciencia se han obtenido así. Pero ¿no pueden haber surgido de forma fortuita, o por alguna modificación del azar como supone Darwin? Respondo que tres o cuatro métodos independientes de cálculo demuestran que sería ridículo suponer que nuestra ciencia ha surgido así…» (CP 6.476)

La inteligencia, o «aprendizaje según la experiencia», abarca una amplia gama de capacidades cognitivas, incluidas las experiencias sensoriales, la imaginación, la memoria, así como los instintos y las emociones, que contribuyen a nuestra comprensión y conocimiento del mundo. La inteligencia está interconectada con la racionalidad, o el razonamiento, el proceso lógico que proporciona una vía para determinar la veracidad de nuestras creencias sobre el mundo y crea hábitos para futuras decisiones y acciones.

Creo que Peirce estaría de acuerdo con Hawkins en que Chat GPT y similares no son máquinas inteligentes, ya que la información que defienden es el resultado de un sofisticado proceso de reconocimiento de patrones y no se recopila a través de la experiencia real en el mundo, que según Peirce es el resultado de la percepción sensorial, la conciencia del mundo «exterior», y no solo de la inferencia deductiva, sino también de la abductiva e inductiva, que proporcionan pruebas sobre la veracidad de nuestras afirmaciones y crean hábitos útiles para acciones futuras.

En un comentario al margen de una reseña de 1894 sobre una nueva traducción de la Ética de Spinoza, mientras participaba en un debate sobre el desarrollo de los teoremas matemáticos, Peirce considera que «sin duda, algunos resultados matemáticos podrían haberse obtenido con la máquina analítica de Babbage» y se pregunta: «pero ¿alguien ha supuesto alguna vez que [las matemáticas] podrían realmente avanzar con una máquina así? Incluso el razonamiento silogístico en sus variedades más elevadas, tal y como aparece en la lógica de las relaciones, *requiere un acto vivo de elección basado en el discernimiento, más allá de las capacidades de cualquier máquina concebible* [énfasis mío]; y esto refuta suficientemente la idea de que el hombre es un mero autómata mecánico dotado de una conciencia ociosa» (CN 2:85).[22] Es algo así como este «acto vivo de elección basado en el discernimiento» lo que interviene en el «fenómeno enormemente complejo» de la mentalidad humana que, según insiste Susan Haack, le falta a A. Roe Bot y que, por lo tanto, le impide firmar legalmente un testamento, como afirma al principio de este artículo.[23]

Conclusión

Hawkins no cree que las máquinas de IA actuales sean verdaderamente inteligentes porque no tienen el mismo sistema de aprendizaje que el cerebro humano, un sistema sensoriomotor que explora y construye modelos del mundo a través del

[22] Peirce se refiere a tres tipos de conciencia: «[1] la conciencia [es] la sensación de un instante… [2] hay una conciencia dual… El pasado es el mundo interior, el presente el mundo exterior… con los fenómenos del error y la ignorancia, que nos obligan a reflexionar… Esta conciencia nos proporciona todos nuestros hechos. Es lo que los convierte en hechos. [3] todo el mundo de las relaciones triádicas, el pensamiento. Somos conscientes de ello» (CP 8.283). También se refiere a la relación entre el primer tipo y la racionalidad: «La conciencia, la sensación del instante que pasa, no tiene, como tal, cabida para la racionalidad. La idea de que la lógica tiene algo que ver con ella es una falacia estrechamente relacionada con el hedonismo en la ética» (CP 2.66).

[23] No hay espacio para entrar en detalles aquí, pero Haack enfatiza la importancia del papel de los artefactos humanos, específicamente sociales, intelectuales e imaginativos, en cualquier explicación de la mentalidad humana.

movimiento y la creación de marcos de referencia. Sí cree que el mecanismo básico que se encuentra en el neocórtex o «cerebro nuevo» puede replicarse en un futuro muy próximo para producir máquinas verdaderamente inteligentes (y conscientes) que creen modelos de partes relevantes del mundo. No cree que sea necesario, ni factible, replicar el mecanismo para otros atributos mentales humanos que se originan en el «cerebro antiguo», más complejo, como las intenciones, los objetivos, las emociones, los instintos, etc., y sugiere que programemos y supervisemos los objetivos de las máquinas de forma individual, junto con «medidas de seguridad». Recientemente, Geoffrey Hinton recomienda lo contrario: que dediquemos la investigación a tratar de inculcar instintos maternales en las máquinas de IA para que aprendan a cuidarnos y no acaben destruyéndonos, como él teme.[^25] Peirce y Haack estarían de acuerdo con Hinton en que los instintos son una parte integral de nuestra inteligencia, pero estarían de acuerdo con Hawkins en que no pueden replicarse en una máquina.[24]

Bibliografía

Cooper, Anderson. 2025. «AI expert: "We'll be toast" without changes in AI technology». YouTube. https://www.youtube.com/watch?v=IidpM2DsrBE.

Gent, Edd. 2018. «These tiny robots could be disease-fighting machines inside the body». MACH. https://www.nbcnews.com/mach/science/these-tiny-robots-could-be-disease-fighting-machines-inside-body-ncna861451.

Haack, Susan. 1973. «The Concept of Mind: A Discussion Between Gilbert Ryle and Susan Haack». The Open University. https://www.open.ac.uk/library/digital-archive/program/video:00525_3017.

———. 2009. *Evidence and Inquiry*. Amherst, NY: Prometheus Books.

———. 2018. «Brave New World: Nature, Culture, and the Limits of Reductionism». En *Explaining the Mind*, editado por Bartosz Brozek et al. Krakow: Copernicus Center Press.

———. 2022a. «That Ryle Interview, in Retrospect». Academia. https://www.academia.edu/81027914/That_Ryle_Interview_in_Retrospect.

———. 2022b. «What a World! The Pluralistic Universe of Innocent Realism». En *Facing the Future, Facing the Screen*, editado por Kristof Niyiri. Budapest: Hungarian Academy of Sciences.

Hawkins, Jeff. 2021. *A Thousand Brains: A New Theory of Intelligence*. New York, NY: Basic Books.

Hinton, Jeffrey. 2025. «Will AI outsmart human intelligence?» YouTube. https://www.youtube.com/watch?v=IkdziSLYzHw.

Migotti, Mark. 2024. «Why study Logic?» En *The Oxford Handbook of Charles Sanders Peirce*, editado por Cornelis de Waal. New York, NY: Oxford University Press.

[24] No se utilizo Chat GPT ni otros LLM en la redacción original de este artículo en inglés. Para la traducción de este artículo, a cargo de los editores de las ponencias, se uso DeepL.

Nubiola, Jaime. 2006. «Peirce en España y anotaciones de Peirce sobre España». *Anthropos Charles Sanders Peirce. Razón e invención del pensamiento pragmatista*, n.º 212: 168-78.

Peirce, Charles. 1975--1987. *Charles Sanders Peirce: Contributions to The Nation*. Editado por Kenneth L. Ketner y James E. Cook. Lubbock: Texas Tech University Press.

———. 1931--1958. *The Collected Papers of Charles Sanders Peirce*. Editado por Charles Hartshorne y Paul Weiss. Cambridge, MA: Harvard University Press.

———. 1887. «Logical Machines». *The American Journal of Psychology* 1 (November): 165-70.

Perez, Carlos. 2025. «Moravec's Paradox from the 4E Cognitive Psychology View». Medium. https://medium.com/intuitionmachine/moravecs-paradox-from-the-4e-cognitive-psychology-view-539359b432fa.

Revuelta, Marta. s. f. «Recension de Clark University, 1889-1899, Decennial Celebration». Online. https://www.unav.es/gep/ClarkUniversity89-99.html.

Rosenwald, Michael. 2025. «Margaret Boden, Philosopher of Artificial Intelligence, Dies at 88». *The New York Times*, agosto. https://www.nytimes.com/2025/08/14/science/margaret-boden-dead.html.

Summerfield, Christopher. 2025. *These Strange New Minds: How AI Learned to Talk and What it Means*. New York, NY: Viking Press.

Wolverton, Troy. 2025. «Palm founder Jeff Hawkins' next big thing is an AI that learns like humans». San Francisco Examiner. https://www.sfexaminer.com/news/technology/why-palm-founder-jeff-hawkins-next-big-thing-could-be-ai/article_b0aede7a-3021-11ef-a127-9f48a3ca5ee8.html.