



Roger Penrose

**La nueva mente
del emperador**

*En torno a la cibernética, la mente y
las leyes de la física*

La nueva mente del emperador es un libro de 1989 escrito por el destacado físico matemático Roger Penrose.

Penrose defiende el argumento de que la consciencia humana no es algorítmica, y por lo tanto no puede ser modelada mediante una máquina de Turing ni computadoras digitales. El autor señala como hipótesis que la mecánica cuántica juega un papel esencial en la comprensión de la consciencia. A su vez, el colapso de la función de onda cuántica jugaría un importante papel en el funcionamiento del cerebro.

El libro abarca varios conceptos relacionados con las ciencias de la computación, las matemáticas y la física cuántica.

Penrose ganó el Premio Aventis en 1990 por este libro.

Dedico este libro a la memoria de mi querida ma-
dre,
quien no vivió para verlo.

NOTA PARA EL LECTOR:

sobre la lectura de las ecuaciones matemáticas

En diferentes partes de este libro he recurrido al uso de ecuaciones matemáticas, desoyendo impertérrito las frecuentes advertencias de que cada una de estas fórmulas reduciría a la mitad el número de lectores. Si usted es una persona que se siente intimidada ante una fórmula (como la mayoría de la gente), entonces le recomiendo un método que yo mismo uso cuando se presenta una de estas fórmulas fastidiosas. El método consiste, más o menos, en pasarla por alto y saltar a la siguiente línea de texto. Bien, no exactamente; es conveniente echar una rápida ojeada a la pobre fórmula, sin tratar de comprenderla del todo, y luego seguir adelante. Algún tiempo después, y armados con nueva confianza, podemos volver a la fórmula olvidada y tratar de captar alguna de sus características más sobresalientes. El propio texto puede servir de ayuda para saber qué es lo importante y qué puede ser pasado por alto sin problemas. Si no lo consigue, entonces prescinda de la fórmula, por completo y sin remordimientos.

AGRADECIMIENTOS

Muchas personas me han ayudado, de una u otra forma, a escribir este libro, y debo darles las gracias. En particular, a los defensores de la IA fuerte (especialmente a los que intervinieron en un programa de la BBC TV que tuve ocasión de presenciar), quienes al expresar opiniones tan radicales me incitaron, hace ya varios años, a embarcarme en este proyecto. (Pese a todo, me temo que si entonces hubiera sabido el esfuerzo que su escritura me iba a exigir, no lo hubiera empezado).

También quiero agradecer a Toby Bailey, David Deutsch (quien también fue de gran ayuda en la comprobación de las especificaciones de mi máquina de Turing), Stuart Hampshire, Jim Hartle, Lane Hughston, Angus McIntyre, Mary Jane Mowat, Tristan Needham, Ted Newman, Eric Penrose, Toby Penrose, Wolfgang Rindler, Engelbert Schücking y Dennis Sciama, quienes revisaron versiones de pequeñas partes del manuscrito y me hicieron muchas sugerencias valiosas para mejorarlo.

Merece un reconocimiento especial la ayuda de Christopher Penrose, con la información detallada respecto al conjunto de Mandelbrot, así como la de Jonathan Penrose, por su valiosa información sobre la computadora que juega ajedrez. Muchas gracias también a Colin Blakemore, Erich Harth y David Hubel por leer y revisar el capítulo IX, que concierne a un tema en el que sinceramente no soy un experto aunque, como sucede con toda la gente que acabo de mencionar, ellos no son en absoluto responsables de los errores que puedan haber quedado. Agradezco a la NSF su

ayuda mediante los contratos DMS 84-05644, DMS 86-06488 y PHY 8612424. Asimismo he contraído una gran deuda con Martin Gardner por su extrema generosidad al escribir el prefacio de este trabajo, y también por sus comentarios. De forma especial, quiero agradecer a mi querida Vanessa por sus atentas y detalladas críticas en varios capítulos, por su invaluable asistencia en las referencias y, lo más importante, por soportarme cuando era insoportable, por su cariño y apoyo donde y cuando era vital.

Procedencia de las imágenes

Los editores agradecen el permiso para reproducir las ilustraciones que se citan:

- Las figuras IV.6 y IV.9 proceden de D. A. Klarner (ed.), *The mathematical Gardner* (Wadsworth International, 1981).
- La figura IV.7 procede de B. Grünbaum y G. C. Shephard, *Tilings and patterns* (W. H. Freeman, 1987). Copyright © 1987 por W. H. Freeman and Company. Utilizada con su permiso.
- La figura IV.10 procede de K. Chandrasekharan, *Hermann Weyl 1885-1985* (Springer, 1986).
- Las figs, IV.11 y X.3 proceden de Pentaplexity: a class of non-periodic tilings of the plane. *The Mathematical Intelligencer*, 2, 32-7 (Springer, 1979).
- La figura IV.12 procede de H. S. M. Coxeter, M. Emmer, R. Penrose y M. L. Teuber (eds.) y M. C. Escher: *Art and science* (North Holland, 1986).
- La figura V.2 © 1989 M. C. Escher Heirs/Cordon Art-Baarn-Holland.
- La figura X.4 procede de *Journal of Materials Research*, 2, 1-4 (Materials Research Society, 1987).

Todas las demás ilustraciones son del autor.

PREFACIO

por MARTIN GARDNER

Para muchos matemáticos y físicos célebres resulta difícil, si no imposible, escribir un libro que pueda ser entendido por los profanos. Hasta hoy se podría haber pensado que Roger Penrose, uno de los físico-matemáticos más eruditos y creativos del mundo, pertenecía a esta clase. Aunque quienes habíamos leído sus artículos y conferencias de divulgación teníamos otra opinión. Aun así, fue una deliciosa sorpresa descubrir que Penrose había robado tiempo a sus ocupaciones para producir un libro maravilloso destinado al profano. Creo que este libro pronto será clásico.

Aunque los capítulos del libro de Penrose recorren la teoría de la relatividad, la mecánica cuántica y la cosmología, su interés principal radica en lo que los filósofos llaman el «problema mente-cuerpo». Durante décadas los defensores de la «IA (Inteligencia Artificial) fuerte» han intentado convencernos de que sólo es cuestión de uno o dos siglos (algunos hablan incluso de cincuenta años), para que las computadoras electrónicas hagan todo lo que la mente humana puede hacer. Estimulados por lecturas juveniles de ciencia-ficción y convencidos de que nuestras mentes son simplemente «computadoras hechas de carne» (como Marvin Minsky dijo en cierta ocasión), dan por supuesto que el placer y el dolor, el gusto por la belleza, el sentido del humor, la consciencia y el libre albedrío son cualidades que emergerán de modo natural cuando el comportamiento al-

gorítmico de los robots electrónicos llegue a ser suficientemente complejo.

Algunos filósofos de la ciencia (en particular John Searle, cuyo famoso experimento mental de la habitación china discute Penrose en detalle) están en abierto desacuerdo. Para ellos una computadora no es esencialmente diferente de las calculadoras mecánicas que funcionan con ruedas, palancas o cualquier otro mecanismo que transmita señales. (Se puede construir una computadora a base de ruedas que giran o agua que se mueva por tuberías). Puesto que la electricidad viaja por los cables conductores mucho más rápido que otras formas de energía (excepto la luz), también puede jugar con los símbolos más rápidamente que las calculadoras mecánicas, y realizar así tareas de enorme complejidad. Pero ¿«comprende» una computadora electrónica lo que está haciendo en una medida superior a la «comprensión» de la que es capaz un ábaco? Las computadoras juegan ahora al ajedrez como un gran maestro. ¿«Comprenden» su juego mejor de lo que lo hace la máquina de jugar a tres en raya que en cierta ocasión construyeron unos desguazadores de computadoras con piezas de chatarra?

Este libro es el ataque más poderoso que se haya escrito contra la IA fuerte. Durante los últimos siglos se han levantado objeciones contra el alegato reduccionista de que la mente es una máquina que funciona según las conocidas leyes de la física, pero la ofensiva de Penrose es más convincente, puesto que hace uso de información de la que no disponían los escritores anteriores. En el libro, Penrose se revela como algo más que un físico-matemático: es también un filósofo de primera línea, que no teme abordar problemas que sus contemporáneos despachan considerándolos sin sentido.

Penrose tiene también el valor de sostener, frente al creciente rechazo de un pequeño grupo de físicos, un vigoroso realismo. No sólo el Universo «está ahí», sino que la ver-

dad matemática tiene también sus propias y misteriosas independencia e intemporalidad. Como Newton y Einstein, Penrose tiene un profundo sentido de humildad y respeto tanto hacia el mundo físico como hacia el ámbito platónico de la matemática pura. Al famoso especialista en teoría de números Paul Erdős le gusta hablar del «libro de Dios» en el que están registradas las demostraciones más notables. A los matemáticos se les permite de cuando en cuando echar una ojeada a alguna página. Penrose cree que cuando un físico o un matemático experimenta una repentina ¡eureka!, no se trata simplemente de algo «producido por un cálculo complicado»: es que la mente, por un momento, entra en contacto con la verdad objetiva. ¿No sería posible, se pregunta, que el mundo de Platón y el mundo físico (que los físicos han diluido ahora en las matemáticas) fueran realmente uno y el mismo?

Muchas páginas del libro están dedicadas a la famosa estructura de tipo fractal conocida como conjunto de Mandelbrot, por ser Benoît Mandelbrot quien la descubrió. Aunque es autosimilar en sentido estadístico, a medida que sus partes son ampliadas, su estructura con infinitas circunvoluciones cambia de manera impredecible. Penrose encuentra incomprensible (igual que yo) que nadie pueda suponer que esta exótica estructura no «esté ahí» igual que lo está el monte Everest, y pueda ser explorada de la misma forma que se explora una selva.

Penrose forma parte del cada vez mayor grupo de físicos que piensan que Einstein no era ni tan obstinado ni tan confuso cuando afirmaba que una «voz interior» le decía que la mecánica cuántica estaba incompleta. Para apoyar esta afirmación, Penrose lleva al lector por un fascinante recorrido a través de temas como los números complejos, las máquinas de Turing, la teoría de la complejidad, las desconcertantes paradojas de la mecánica cuántica, los sistemas formales, la indecidibilidad de Gödel, los espacios fase, los espacios de Hilbert, los agujeros negros, los agujero-

ros blancos, la radiación de Hawking, la entropía o la estructura del cerebro, y tantea otros temas que están en el centro de las especulaciones actuales. ¿Tienen los perros y los gatos «consciencia» de sí mismos? ¿Es posible, en teoría, para una máquina que transmite materia, transferir a una persona de un lugar a otro de la misma manera en que eran transmitidos o recibidos los astronautas de la serie de televisión *Star Trek*? ¿Cómo ayudó a la supervivencia el que la evolución haya producido la consciencia? ¿Existe un nivel más allá de la mecánica cuántica en el que la dirección del tiempo y la distinción entre izquierda y derecha estén indisolublemente asociados? ¿Son las leyes de la mecánica cuántica —o quizás otras leyes aún más profundas— esenciales para la actuación de la mente?

La respuesta de Penrose a las dos últimas preguntas es afirmativa. Su famosa teoría de los twistors —objetos geométricos abstractos que operan en un espacio complejo multidimensional que subyace bajo el espacio-tiempo— es demasiado técnica para ser incluida en este libro. Ellos representan los esfuerzos de Penrose durante dos décadas para sondear una región más profunda que la de los campos y las partículas en la mecánica cuántica. Al clasificar las teorías en cuatro categorías: extraordinarias, útiles, provisionales y erróneas, Penrose coloca modestamente su teoría de los twistors en la clase de las provisionales, junto con la de las supercuerdas u otros grandes esquemas unificadores que hoy son fuertemente debatidos.

Penrose es, desde 1973, el catedrático Rouse Ball de Matemáticas en la Universidad de Oxford. El título es apropiado ya que W. W. Rouse Ball no sólo fue un notable matemático sino también un mago aficionado, con un interés tan apasionado por las matemáticas recreativas que escribió una obra clásica en este campo: *Mathematical Recreations and Essays*. Penrose comparte el entusiasmo de Ball por el juego. En su juventud descubrió un «objeto imposible» llamado «tribar». (Un objeto imposible es el dibujo de

una figura sólida que no puede existir ya que incorpora elementos contradictorios). Él y su padre Lionel, genetista, convirtieron el tribar en la Escalera de Penrose, una estructura que Maurits Escher utilizó en dos famosas litografías «Ascenso y Descenso» y «Cascada». Un día en que Penrose estaba tumbado en la cama imaginó, en lo que él llamó un «arrebato de locura», un objeto imposible en un espacio tradimensional. Es algo, decía, que si se le mostrase a una criatura del espacio de cuatro dimensiones le haría exclamar «¿qué es esto? ¡Dios mío!».

Durante los años sesenta, mientras trabajaba en cosmología con su amigo Stephen Hawking, Penrose hizo el que tal vez sea su descubrimiento más conocido. Si la teoría de la relatividad es válida «hasta el final», en todo agujero negro debe haber una singularidad en la que ya no sean aplicables las leyes de la física. Incluso este resultado ha sido eclipsado últimamente por la construcción que él mismo hizo de dos formas que embaldosan el plano, a la manera de la teselación de Escher, pero que sólo pueden hacerlo en forma no periódica. (Encontrará una discusión de estas sorprendentes formas en mi libro *Penrose Tiles to Trapdoor Ciphers*). Penrose las inventó, o más bien las descubrió, sin esperar que fueran de utilidad. Para asombro de todos resultó que las formas tridimensionales de sus baldosas pueden subyacer bajo un extraño y nuevo tipo de materia. El estudio de estos «cuasicristales» es hoy en día una de las áreas de investigación más activas dentro de la cristalografía. Es también el ejemplo más espectacular en los tiempos modernos de cómo las matemáticas lúdicas pueden tener aplicaciones no previstas.

Los resultados de Penrose en matemáticas y física —y sólo he mencionado una pequeña parte— surgen de una permanente admiración por el misterio y por la belleza del ser. Su voz interior le dice que la mente humana es algo más que una simple colección de minúsculos cables e interruptores. El Adam de sus prólogo y epílogo es en parte el

símbolo del despertar de la consciencia en la lenta evolución de la vida sensible. Para mí, Penrose es también el niño sentado en la tercera fila, detrás de las vacas sagradas de la IA, y que se atreve a sugerir que el emperador de la IA fuerte va desnudo. Aunque sus opiniones estén salpicadas de humor, ésta no es materia de risa.

PRÓLOGO

Hay una numerosa concurrencia en el Gran Auditorio para asistir a la presentación de la nueva computadora «Ultronic». El presidente Polho acaba de concluir su discurso de apertura y se alegra de ello: no se siente a gusto en tales ocasiones y no sabe nada de computadoras, salvo que ésta le va a ahorrar mucho tiempo. Los fabricantes le han asegurado que, entre sus muchos cometidos, será capaz de asumir todas las delicadas decisiones de Estado tan fastidiosas para él. Mejor que así sea, considerando la cuantiosa suma que se ha invertido en ello. Se ve ya disfrutando de muchas horas libres para jugar al golf en su magnífico campo privado, una de las pocas áreas verdes extensas que quedan en su pequeño país.

Adam se sentía privilegiado de contarse entre los asistentes a la ceremonia de inauguración. Se sentó en la tercera fila. Dos filas más adelante de él estaba su madre, tecnócrata que había intervenido en el diseño de Ultronic. Casualmente su padre también estaba allí —en el fondo de la sala—, completamente rodeado de guardias de seguridad. En el último minuto el padre de Adam había tratado de hacer estallar la computadora. Él mismo se había encomendado esta misión, autonombrándose «espíritu conductor» de un pequeño grupo de activistas: el Gran Consejo para la Consciencia Psíquica. Por supuesto, él con todos sus explosivos habían sido inmediatamente detectados por los numerosos sensores electrónicos y químicos. Una pequeña

parte de su castigo consistiría en ser testigo de la ceremonia de inauguración.

Adam no sentía especial aprecio por sus padres. Quizá no necesitaba tales sentimientos. Durante sus trece años había sido criado casi exclusivamente por computadoras rodeado de todas las comodidades. Podía tener todo lo que quisiera sin más que apretar un botón: comida, bebida, compañía y entretenimiento; también información sobre cualquier cosa que le interesara, siempre ilustrada con coloridas y atractivas ilustraciones. La alta posición de su madre había hecho posible todo esto.

El diseñador en jefe estaba llegando al final de su discurso: «... tiene más de 10^{17} unidades lógicas. ¡Más que el número total de neuronas que reúnen todos los cerebros de todas las personas en todo el país! Su inteligencia será inimaginable. Afortunadamente, sin embargo, no necesitamos imaginarla. Dentro de un instante todos nosotros tendremos el Privilegio de ser testigos de primera mano de su inteligencia: ¡pido a la respetable primera dama de nuestro gran país, la señora Isabella Polho, que conecte el interruptor que activa nuestra fantástica computadora Ultronic!».

La esposa del presidente avanzó. Un poco nerviosa, y con cierta torpeza, cerró el interruptor. Se produjo un gran silencio y un casi imperceptible parpadeo de las luces cuando las 10^{17} unidades lógicas se activaron. Todos esperaban, sin saber muy bien el qué. «Bien, ¿hay alguien en la audiencia que quiera dirigirse a nuestro nuevo Sistema de Cómputo Ultronic para plantearle la primera pregunta?», interrogó el diseñador en jefe. Nadie se atrevía, temerosos de parecer estúpidos ante la multitud, y ante la nueva omnipresencia. Se hizo el silencio. «Sin duda hay alguien», suplicó. Pero todos tenían miedo, aprensivos frente a la nueva y todopoderosa consciencia. Pero Adam no sentía el mismo respeto, por el hecho de haber crecido entre computadoras. Casi sabía lo que se sentiría ser una computadora, o